



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/87, A61K 48/00 C12N 15/34, 15/86, 7/04 C12N 15/62, 15/13		A1	(11) International Publication Number: WO 94/10323 (43) International Publication Date: 11 May 1994.(11.05.94)
(21) International Application Number: PCT/GB93/02267		(74) Agent: BASSETT, Richard, S.; Eric Potter Clarkson, St. Mary's Court, St. Mary's Gate, Nottingham NG1 1LE (GB).	
(22) International Filing Date: 4 November 1993 (04.11.93)		(81) Designated States: GB, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority data: 9223084.6 4 November 1992 (04.11.92) GB		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
<p>(71) Applicant (for all designated States except US): IMPERIAL CANCER RESEARCH TECHNOLOGY LIMITED [GB/GB]; Sardinia House, Sardinia Street, London WC2A 3NL (GB).</p> <p>(72) Inventor; and</p> <p>(75) Inventors/Applicants (for US only): SPOONER, Robert, Anthony [GB/GB]; EPENETOS, Agamemnon, Antoniou [GB/GB]; ICRF Oncology Group, Department of Clinical Oncology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London WC12 0HS (GB).</p>			
(54) Title: VIRUS WITH MODIFIED BINDING MOIETY SPECIFIC FOR THE TARGET CELLS			
<p>(57) Abstract.</p> <p>A virus, or virus-like particle, derived from a virus or virus-like particle having a receptor for a host cell comprising a modified binding specificity conferred by a binding moiety allowing the virus or virus-like particle to bind to a target cell characterized in that the said host cell receptor is modified or absent so that the virus or virus-like particle is substantially incapable of binding the said host cell are disclosed. An adenovirus or influenza virus or vaccinia virus, or a replication defective derivative of any of these characterized in that the virus has a modified binding specificity conferred by a binding moiety allowing the virus to bind to a target cell are disclosed. Suitable binding moieties include monoclonal antibodies, ScFVs, dAbs and minimal recognition units. The use of at least some of these as delivery vehicles for genes to target cells in the fields of gene therapy and cancer treatment are disclosed.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BP	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
LI	Lithuania	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KB	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LK	Lithuania	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CZ	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## VIRUS WITH MODIFIED BINDING MOIETY SPECIFIC FOR THE TARGET CELLS.

The present invention relates to delivery vehicles for genes to target cells, especially in the fields of gene therapy and cancer treatment.

5

The delivery of genes to target cells, especially those within the mammalian body, has many uses, for example in the fields of gene therapy, cancer treatment and in areas of genetic manipulation still to be discovered. The gene to be delivered may encode a molecule, such as a 10 protein or RNA, which is cytotoxic to the target cell, or it may encode a functional copy of a gene that is defective in the target cell. In this latter case the product of the aforementioned functional copy of the gene will replace that of the defective copy, and the target cell will be able to perform its proper function.

15

The use of viruses, or virus-like particles, to deliver genes for gene therapy and cancer treatment has been disclosed.

However, in most cases the targeting of the virus or virus-like particles 20 containing the desired gene to the cell has relied on the natural host-virus specificity or on local application of the virus to the cells to be targeted, for example direct application of viruses to lung cells by inhalation.

The human adenovirus 5 (Ad5) genome consists of a double-stranded 25 linear DNA molecule of 36 kilo-basepair. The virus replication cycle has two phases: an early phase, during which four transcriptional units E1, E2, E3, and E4 are expressed, and a late phase occurring after the onset of viral DNA synthesis when late transcripts are expressed from the major late promoter (MLP). These late messages encode most of the viral 30 structural proteins. E1, E2, and E4 gene products of human adenoviruses

(Ads) are involved in transcriptional activation, cell transformation, and viral DNA replication as well as other viral functions, and are essential for viral growth. In contrast, E3 gene products are not required for viral replication in cultured cells or for acute lung infection of cotton rats, but 5 appear to be involved in evading immune surveillance *in vivo*.

By "virus-like particle" we mean a nucleoprotein particle containing a core of nucleic acid surrounded by protein which (i) is not infective and (ii) can only be propagated in a suitable cell system following 10 transformation by its nucleic acid. Thus a virus-like particle of mammalian origin may be propagated in *Saccharomyces cerevisiae* or in insect cells *via* a baculovirus expression system.

15 The modification of coat proteins of filamentous bacteriophages (bacterial viruses), such as M13 and fd, so as to generate novel binding properties, has been disclosed in Cwirla *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6378-6382 and Scott & Smith (1990) *Science* **249**, 386-390.

20 It has previously been suggested that retroelement particles, including retroviral vectors, may be modified to target specific cells, for example see Kingsman *et al* (1991) *Tibtech* **9**, 303-309.

25 Russell *et al* (1993) *Nucl. Acids Res.* **21**, 1081-1085, published after the priority date for this application but before the filing date discloses retroviral vectors displaying functional antibody fragments and suggests that, in principle, the display of antibody fragments on the surface of recombinant retroviral particles could be used to target virus to cells for gene delivery. However, it is not known whether a retrovirus can be assembled in which all the subunits of the viral envelope protein are fused 30 to antibody, and if so whether the virus would infect cells.

NIP-derivatised human cells were tested as a method for targeted gene delivery, but became permissive for both modified (displaying an anti-NIP antibody) and unmodified ecotropic viral particles. NIP is 4-hydroxy-3-iodo-5-nitrophenoxyacetic acid.

5

Michael *et al* (1993) *J. Biol. Chem.* 268, 6866-6869, published after the priority date of this application but before the filing date, describes molecular conjugates between adenovirus and a vector system comprising two linked domains, a DNA binding domain and a ligand domain. In this 10 configuration, however, it is stated that the viral moiety functions in the capacity of both an alternate ligand domain of the conjugate and, since an additional ligand has been introduced into the conjugate design, the potential for cell-specific targeting is undermined.

15 Curiel *et al* (1992) *Human Gene Therapy* 3, 147-154 describes adenoviruses wherein a foreign epitope was introduced into the hexon protein and polylysine-antibody complexed DNA was attached to adenovirus by virtue of the antibody binding the foreign epitope on the hexon. Foreign DNA is transferred bound to the exterior of the virion.

20 The above-mentioned viruses and virus-like particles may be able to target cells using the binding moiety displayed on their surface but they can also still target their natural host cells.

25

We have now devised new viruses and virus-like particles at least some of which can bind the target cell with high specificity and may deliver genetic material to the target cell; at least some of the viruses and virus-like particles may bind and deliver genetic material to the target cell without substantially binding to the natural host cell of the virus.

30

One aspect of the present invention provides a virus, or virus-like particle, derived from a virus or virus-like particle having a receptor for a host cell comprising a modified binding specificity conferred by a binding moiety allowing the virus or virus-like particle to bind to a target cell

5 characterised in that the said host cell receptor is modified or absent so that the virus or virus-like particle is substantially incapable of binding the said host cell.

By "substantially incapable of binding its host cell" we mean that the 10 modified virus has no more than 1% of the binding affinity of the unmodified virus for the host cell.

In general, the binding specificity of a natural virus or virus-like particle is conferred by the specific interaction between a receptor-like molecule 15 expressed on the surface of the virus or virus-like particle and a cognate receptor-like molecule expressed on the surface of its host cell. The invention provides a beneficial modification of the binding specificity, so that the virus or virus-like particle can bind to a different specific target cell.

20 The introduction of the modified binding moiety may be such as to achieve the said removal of the native binding specificity.

A second aspect of the invention comprises an adenovirus or influenza 25 virus or vaccinia virus, or a replication-defective derivative of any of these, characterised in that the virus has a modified binding specificity conferred by a binding moiety allowing the virus to bind to a target cell.

30 By "binding moiety" we mean a molecule that is exposed on the surface of the virus or virus-like particle which is able to bind to a molecule on

the target cell. The "binding moiety" may be a molecule on the virus or virus-like particle modified in such a way that its binding specificity is changed, or it may be a molecule added to, and exposed on the surface of, the virus or virus-like particle to provide a new binding specificity.

It is preferred if the binding moiety is external to the receptor for its host cell of the naive, unmodified virus.

It is further preferred if the binding moiety is joined or fused to the virus  
10 or virus-like particles directly or indirectly by a spacer group.

By "host cell" we mean the cell that an unmodified, naive virus can bind to using its receptor-like molecule and the cognate receptor-like molecule on the cell. By "target cell" we mean the cell that the modified virus can  
15 bind to using its binding moiety. In some circumstances in the context of the second aspect of the invention, such as when the binding moiety recognises an entity on the host cell which is not the cognate receptor-like molecule, then the host cell may be the target cell.

20 The virus or virus-like particle may be a bacteriophage and the target cell a bacterium in which case the invention may find uses in the treatment of bacterial infections.

In a preferred embodiment of the invention the target cell is eukaryotic.  
25 The eukaryotic cell may be a yeast cell and the virus or virus-like particle may be useful in the medical field in treating yeast infections such as athlete's foot or *Candida* infection but it is preferred that the eukaryotic cell is mammalian, and it is expected that the invention will find uses in the areas of gene therapy and cancer treatment.

In preferred embodiments of the first aspect of the invention the virus or virus-like particle is adenovirus or influenza virus or a pox-virus such as vaccinia.

- 5 It is also preferred that the virus or virus-like particle is "replication-defective". By "replication defective" we mean a virus whose genetic material has been manipulated so that it cannot divide or proliferate in the cell it infects.
- 10 The binding moiety of the virus or virus-like particle of the invention provides the target cell binding specificity. Any cell-binding protein or peptide or carbohydrate or lipid may be useful for targeting the virus or virus-like particle to the cell. For example, short linear stretches of amino acids, such as those constituting a peptide hormone, are useful, as are
- 15 domains of polypeptides that can fold independently into a structure that can bind to the target cell.

In one preferred embodiment the binding moiety has the property of any one of a monoclonal antibody, ScFv (single chain Fv fragment), a dAb (single domain antibody) or a minimal recognition unit of an antibody.

The binding site on the target cell may be a target cell-specific antigen. Such antigens are listed in Table 1. Other binding moieties, targets on cells, and diseases which could usefully be treated using reagents delivered by the modified viruses or virus-like particles are given in Table 2.

Table 1

1. <u>Tumour Associated Antigens</u>			
	<u>Antigen</u>	<u>Antibody</u>	<u>Existing Uses</u>
5	Carcino-embryonic Antigen	{C46 (Amersham) 85A12 (Unipath)}	Imaging & Therapy of colon/rectum tumours.
	Placental Alkaline Phosphatase	HI7E2 (ICRF, Travers & Bodmer)	Imaging & Therapy of testicular and ovarian cancers.
	Pan Carcinoma	NR-LU-10 (NeoRx Corporation)	Imaging & Therapy of various carcinomas incl. small cell lung cancer.
10	Polymorphic Epithelial Mucin (Human milk fat globule)	HMFG1 (Taylor-Papadimitriou, ICRF)	Imaging & Therapy of ovarian cancer, pleural effusions.

$\beta$ -human Chorionic Gonadotropin W14 Targeting of enzyme (CPG2) to human xenograft choriocarcinoma in nude mice. (Searle *et al* (1981) *Br. J. Cancer* 44, 137-144)

A Carbohydrate on L6 (IgG2a)<sup>1</sup> Targeting of alkaline phosphatase. (Senter *et al* (1988) *P.N.A.S.* 85, 4842-4846

Human Carcinomas

5 CD20 Antigen on B Lymphoma (normal and neoplastic) 1F5 (IgG2a)<sup>2</sup> Targeting of alkaline phosphatase. (Senter *et al* (1988) *P.N.A.S.* 85, 4842-4846

<sup>1</sup>Hellström *et al* (1986) *Cancer Res.* 46, 3917-3923

<sup>2</sup>Clarke *et al* (1985) *P.N.A.S.* 82, 1766-1770

10

Other antigens include alphafoetoprotein, Ca-125 and prostate specific antigen.

## 2. Immune Cell Antigens

15 Pan T Lymphocyte Surface Antigen OKT-3 (Ortho) (CD3) As anti-rejection therapy for kidney transplants.

	B-lymphocyte Surface Antigen (CD22)	RFB4 (Janossy, Royal Free Hospital)	Immunotoxin therapy of B cell lymphoma.
5	Pan T lymphocyte Surface Antigen (CD5)	H65 (Bodmer, Knowles ICRF, Licensed to Xoma Corp., USA)	Immunotoxin treatment of Acute Graft versus Host disease, Rheumatoid Arthritis.

3. Infectious Agent-Related Antigens

	Mumps virus-related	Anti-mumps polyclonal antibody	Antibody conjugated to Diphtheria toxin for treatment of mumps.
10	Hepatitis B Surface Antigen	Anti HBs Ag	Immunotoxin against Hepatoma.

Table 2: Binding moieties for tumour-specific targets and tumour  
associated antigens

Target	Binding moiety	Disease
5 Truncated EGFR Idiotypes EGFR ( <i>c-erbB1</i> ) <i>c-erbB2</i>	anti-EGFR mAb anti-id mAbs EGF, TGF $\alpha$ anti-EGFR mAb	Gliomas B-cell lymphomas Breast cancer
	mAbs	Breast cancer
	IL-2	Lymphomas
	anti-Tac mAb	and leukaemias
10 IL-4 receptor IL-6 receptor	IL-4	Lymphomas
	IL-6	and leukaemias
	$\alpha$ -MSH	Lymphomas
15 MSH (melanocyte-stimulating hormone) receptor Transferrin receptor (TR)	Transferrin	and leukaemias
	anti-TR mAb	Melanomas
20 gp95/gp97 p-glycoprotein cells cluster-1 antigen (N-CAM) cluster-w4	mAbs	Gliomas
	mAbs	Melanomas
	mAbs	drug-resistant
	mAbs	Small cell lung carcinomas
25 cluster-5A cluster-6 (LeY)	mAbs	Small cell lung carcinomas
	mAbs	carcinomas
	mAbs	Small cell lung carcinomas
30 PLAP (placental alkaline phosphatase)	mAbs	Some seminomas
	mAbs	Some ovarian; some non-small cell lung cancer
35 CA-125 ESA (epithelial specific antigen) CD 19, 22, 37 250 kDa proteoglycan	mAbs	Lung, ovarian carcinoma
	mAbs	B-cell lymphoma
	mAbs	Melanoma
	mAbs	Breast cancer
40 p55 TCR-IgH fusion	mAbs	Childhood T-cell leukaemia
	mAbs	Gastric and colon tumours
	mAbs	
45 Blood gp A antigen (in B or O individuals)		

The binding moiety may be a monoclonal antibody. Monoclonal antibodies which will bind to many of these antigens are already known but in any case, with today's techniques in relation to monoclonal antibody technology, antibodies can be prepared to most antigens. The binding

5 moiety may be a part of an antibody (for example a Fab fragment) or a synthetic antibody fragment (for example, ScFv). Suitable monoclonal antibodies to selected antigens may be prepared by known techniques, for example those disclosed in "*Monoclonal Antibodies: A manual of techniques*", H Zola (CRC Press, 1988) and in "*Monoclonal Hybridoma Antibodies: Techniques and Applications*", J G R Hurrell (CRC Press, 10 1982).

Suitably prepared non-human antibodies can be "humanized" in known ways, for example by inserting the CDR regions of mouse antibodies into 15 the framework of human antibodies.

The variable heavy ( $V_H$ ) and variable light ( $V_L$ ) domains of the antibody are involved in antigen recognition, a fact first recognised by early protease digestion experiments. Further confirmation was found by 20 "humanization" of rodent antibodies. Variable domains of rodent origin may be fused to constant domains of human origin such that the resultant antibody retains the antigenic specificity of the rodent parental antibody (Morrison *et al* (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6851-6855).

25 That antigenic specificity is conferred by variable domains and is independent of the constant domains is known from experiments involving the bacterial expression of antibody fragments, all containing one or more variable domains. These molecules include Fab-like molecules (Better *et al* (1988) *Science* **240**, 1041); Fv molecules (Skeirra *et al* (1988) *Science* **240**, 1038); ScFv molecules where the  $V_H$  and  $V_L$  partner domains are 30

linked via a flexible oligopeptide (Bird *et al* (1988) *Science* 242, 423; Huston *et al* (1988) *Proc. Natl. Acad. Sci. USA* 85, 5879) and dAbs comprising isolated V domains (Ward *et al* (1989) *Nature* 341, 544). A general review of the techniques involved in the synthesis of antibody

5 fragments which retain their specific binding sites is to be found in Winter & Milstein (1991) *Nature* 349, 293-299.

By "ScFv molecules" we mean molecules wherein the V<sub>H</sub> and V<sub>L</sub> partner domains are linked via a flexible oligopeptide.

10

It may be advantageous to use antibody fragments, rather than whole antibodies. Effector functions of whole antibodies, such as complement binding, are removed. ScFv and dAb antibody fragments can be expressed as fusions with other polypeptides.

15

Minimal recognition units may be derived from the sequence of one or more of the complementary-determining regions (CDR) of the Fv fragment. Whole antibodies, and F(ab')<sub>2</sub> fragments are "bivalent". By "bivalent" we mean that the said antibodies and F(ab')<sub>2</sub> fragments have

20 two antigen combining sites. In contrast, Fab, Fv, ScFv, dAb fragments and minimal recognition units are monovalent, having only one antigen combining sites.

In a further embodiment the binding moiety is at least part of a ligand of

25 a target cell-specific cell-surface receptor.

It is preferred that the target cell-specific cell-surface receptor is the receptor for human gonadotrophin releasing hormone (GnRH). In this preferred embodiment the binding moiety is GnRH, and its binding

30 specificity is for human cancer cells that express the GnRH receptors on

their surface. Examples of such human cancer cells are prostate, breast and endometrial cancer cells.

It is also preferred that the target cell-specific cell-surface receptor is the  
5 receptor for melanocyte-stimulating hormone (MSH) which is expressed in high number in melanoma cells. In this preferred embodiment the binding moiety is MSH, and its binding specificity is for melanoma cells.

It is also preferred that the target cell-specific cell-surface receptor is the  
10 receptor for somatostatin.

Of course, the receptors for GnRH, MSH and somatostatin may themselves be target cell-specific antigens and may be recognised by binding moieties which have the property of any one of a monoclonal antibody, a ScFv, a dAb or a minimal recognition unit. Thus, although the binding site on the target cell may be a cell-surface receptor it may also act as a target cell-specific cell-surface antigen for recognition by the binding moiety.

20 It will be appreciated by those skilled in the art that binding moieties which are polypeptides may be conveniently made using recombinant DNA techniques. The binding moiety may be fused to a protein on the surface of the virus or virus-like protein as disclosed below or they may be synthesised independently of the virus or virus-like particle, by  
25 expression from a suitable vector in a suitable host and then joined to the virus or virus-like particle as disclosed below.

30 Nucleic acid sequences encoding many of the targeting moieties are known, for example those for peptide hormones, growth factors, cytokines and the like and may readily be found by reference to publicly accessible

nucleotide sequence databases such as EMBL and GenBank. Once the nucleotide sequence is known it is obvious to the person skilled in the art how to make DNA encoding the chosen binding moiety using, for example, chemical DNA synthetic techniques or by using the polymerase chain reaction to amplify the required DNA from genomic DNA or from tissue-specific cDNA.

Many cDNAs encoding peptide hormones, growth factors, cytokines and the like, all of which may be useful as binding moieties, are generally

10 available from, for example British Biotechnology Ltd, Oxford, UK.

It is preferred that when the virus or virus-like particle of the invention binds to its target cell it delivers its nucleic acid to the said target cell, that is the target cell is infected by the virus or virus-like particle. Target 15 cells, especially cancer cells, that are infected in this manner by the virus or virus-like particle may express viral molecules on their surface and may be recognised by the immune system and destroyed. Of course, other cytotoxic functions of the virus may also kill the cell.

20 In one embodiment when the virus or virus-like particle is adenovirus, the E1B gene is substantially deleted or modified so that its gene product no longer interacts with the E1A protein. E1A protein stimulates apoptosis but normally its action is inhibited by E1B. Conveniently, the E1B gene is inactivated by insertion; preferably a cytotoxic gene, as defined below, 25 is inserted at or near the E1B gene.

E1, E3 and a site upstream of E4 may be used as sites for insertion of foreign DNA sequences in the generation of recombinant adenoviruses for example see Berkner and Sharp (1984) *Nucl. Acids Res.* 12, 1925-1941; 30 Chanda *et al* (1990) *Virology* 175, 535-547; Haj-Ahmad and Graham

(1986) *J. Virol.* 57, 267-274; Saito *et al* (1985) *J. Virol.* 54, 711-719; all incorporated herein by reference. Since the upper size limit for DNA molecules that can be packaged into adenovirus particles is approximately 105% of the wild-type genome only about 2 kb of extra DNA can be 5 inserted without compensating deletions of viral DNA. Although E1 is essential for virus replication in cell culture, foreign DNA can be substituted for E1 sequences when the virus is grown in 293 cells which are transformed by Ad5 DNA and constitutively express E1 (Graham *et al* (1977) *J. Gen. Virol.* 36, 59-72, incorporated herein by reference).

10 Several vectors having 1.9 kb deleted from E3 of Ad5 have been constructed without interfering with virus replication in cell culture (reviewed by Graham and Prevec (1992) in "Vaccines: New Approaches to Immunological Problems" R.W. Ellis (Ed.), Butterworth-Heinemann, Boston, MA, pages 364-390, incorporated herein by reference). Such 15 vectors allow for insertion of up to 4 kb of foreign DNA. Recombinant adenoviruses containing inserts in E3 replicate in all Ad-permissive cell lines and a number of adenovirus vectors containing E3 inserts have been shown to express foreign genes efficiently both *in vitro* and *in vivo* (Berkner (1988) *Biotechniques* 6, 616-629; Chanda *et al* (1990) *Virology* 20 175, 535-547; Dewar *et al* (1989) *J. Virol.* 63, 129-136; Graham (1990) *Trends Biotechnol.* 8, 85-87; Graham and Prevec (1992) in "Vaccines: New Approaches to Immunological Problems" R.W. Ellis (Ed.), Butterworth-Heinemann, Boston, MA, pages 364-390; Johnson *et al* (1988) *Virology* 164, 1-14; Lubeck *et al* (1989) *Proc. Natl. Acad. Sci. USA* 86, 6763-6767; McDermott *et al* (1989) *Virology* 169, 244-247; Morin *et al* (1987) *Proc. Natl. Acad. Sci. USA* 84, 4626-4630; Prevec *et al* (1989) *J. Gen. Virol.* 70, 429-434; Prevec *et al* (1990) *J. Inf. Dis.* 161, 27-30; Schneider *et al* (1989) *J. Gen. Virol.* 70, 417-427; Vernon *et al* (1991) *J. Gen. Virol.* 72, 1243-1251; Yuasa *et al* (1991) *J. Gen. Virol.* 30 72, 1927-1934) all incorporated herein by reference.

Substantially replication-defective adenoviruses may be made by creating a deficiency of the E1A protein. Suitably this is achieved by deleting the E1A gene or by making mutations within the E1A gene that prevent expression of the E1A protein. Examples of suitable mutations are 5 deletions within the E1A coding region; nonsense mutations; and frameshift mutations.

In further preference, the virus or virus-like particle is modified further to contain a gene suitable for gene therapy.

10

In one embodiment, the gene encodes a molecule having a directly or indirectly cytotoxic function. By "directly or indirectly" cytotoxic, we mean that the molecule encoded by the gene may itself be toxic (for example ricin; tumour necrosis factor; interleukin-2; interferon-gamma; 15 ribonuclease; deoxyribonuclease; *Pseudomonas exotoxin A*) or it may be metabolised to form a toxic product, or it may act on something else to form a toxic product. The sequence of ricin cDNA is disclosed in Lamb *et al* (1985) *Eur. J. Biochem.* 148, 265-270 incorporated herein by reference.

20

For example, it would be desirable to target a DNA sequence encoding an enzyme using the virus or virus-like particle of the invention, the enzyme being one that converts a relatively non-toxic prodrug to a toxic drug. The enzyme cytosine deaminase converts 5-fluorocytosine (5FC) to 5-fluorouracil (5FU) (Mullen *et al* (1922) *PNAS* 89, 33); the herpes simplex 25 enzyme thymidine kinase sensitises cells to treatment with the antiviral agent ganciclovir (GCV) or aciclovir (Moolten (1986) *Cancer Res.* 46, 5276; Ezzedine *et al* (1991) *New Biol* 3, 608). The cytosine deaminase of any organism, for example *E. coli* or *Saccharomyces cerevisiae*, may 30 be used.

Thus, in a preferred embodiment of the invention, the gene encodes a cytosine deaminase and the patient is concomitantly given 5FC. By "concomitantly", we mean that the 5FC is administered at such a time, in relation to the transformation of the tumour cells, that 5FC is converted 5 into 5FU in the target cells by the cytosine deaminase expressed from the said gene. A dosage of approximately 0.001 to 100.0 mg 5FC/kg body weight/day, preferably 0.1 to 10.0 mg/kg/day is suitable.

Components, such as 5FC, which are converted from a relatively non-10 toxic form into a cytotoxic form by the action of an enzyme are termed "pro-drugs".

Other examples of pro-drug/enzyme combinations include those disclosed by Bagshawe *et al* (WO 88/07378), namely various alkylating agents and 15 the *Pseudomonas* spp. CPG2 enzyme, and those disclosed by Epenetos & Rowlinson-Busza (WO 91/11201), namely cyanogenic pro-drugs (for example amygdalin) and plant-derived  $\beta$ -glucosidases.

Enzymes that are useful in this embodiment of the invention include, but 20 are not limited to, alkaline phosphatase useful for converting phosphate-containing prodrugs into free drugs; arylsulfatase useful for converting sulfate-containing prodrugs into free drugs; cytosine deaminase useful for converting non-toxic 5-fluorocytosine into the anti-cancer drug, 5-fluorouracil; proteases, such as serratia protease, thermolysin, subtilisin, 25 carboxypeptidases and cathepsins (such as cathepsins B and L), that are useful for converting peptide-containing prodrugs into free drugs; D-alanylcarboxypeptidases, useful for converting prodrugs that contain D-amino acid substituents; carbohydrate-cleaving enzymes such as  $\beta$ -galactosidase and neuraminidase useful for converting glycosylated 30 prodrugs into free drugs;  $\beta$ -lactamase useful for converting drugs

derivatized with  $\beta$ -lactams into free drugs; and penicillin amidases, such as penicillin V amidase or penicillin G amidase, useful for converting drugs derivatized at their amine nitrogens with phenoxyacetyl or phenylacetyl groups, respectively, into free drugs. Alternatively, 5 antibodies with enzymatic activity, also known in the art as abzymes, can be used to convert the prodrugs of the invention into free active drugs [see, e.g. R J Massey, *Nature*, 328, pp. 457-458 (1987)].

Similarly, the prodrugs of this invention include, but are not limited to, 10 the above-listed prodrugs, e.g., phosphate-containing prodrugs, thiophosphate-containing prodrugs, sulfate-containing prodrugs, peptide-containing prodrugs, D-amino acid-modified prodrugs, glycosylated prodrugs,  $\beta$ -lactam-containing prodrugs, optionally substituted phenoxyacetamide-containing prodrugs or optionally substituted 15 phenylacetamide-containing prodrugs, 5-fluorocytosine and other 5-fluorouridine prodrugs which can be converted by the enzyme of the conjugate into the more active, cytotoxic free drug. Examples of cytotoxic drugs that can be derivatized into a prodrug form for use in this invention include, but are not limited to, etoposide, teniposide, 20 adriamycin, daunomycin, carminomycin, aminopterin, dactinomycin, mitomycins, cis-platinum and cis-platinum analogues, bleomycins, esperamicins [see U.S. Pat. No. 4,675,187], 5-fluorouracil, melphalan and other related nitrogen mustards.

25 In a further embodiment the gene delivered to the target cell encodes a ribozyme capable of cleaving targeted RNA or DNA. The targeted RNA or DNA to be cleaved may be RNA or DNA which is essential to the function of the cell and cleavage thereof results in cell death or the RNA or DNA to be cleaved may be RNA or DNA which encodes an 30 undesirable protein, for example an oncogene product, and cleavage of

this RNA or DNA may prevent the cell from becoming cancerous.

Ribozymes which may be encoded in the genomes of the viruses or virus-like particles herein disclosed are described in Cech and Herschlag "Site-specific cleavage of single stranded DNA" US 5,180,818; Altman *et al* "Cleavage of targeted RNA by RNase P" US 5,168,053, Cantin *et al* "Ribozyme cleavage of HIV-1 RNA" US 5,149,796; Cech *et al* "RNA ribozyme restriction endoribonucleases and methods", US 5,116,742; Been *et al* "RNA ribozyme polymerases, dephosphorylases, restriction endonucleases and methods; cleaves single-stranded RNA at specific site by transesterification", US 4,987,071, all incorporated herein by reference.

15 In a still further embodiment the gene delivered to the target cell encodes an antisense RNA.

By "antisense RNA" we mean an RNA molecule which hybridises to, and interferes with the expression from a mRNA molecule encoding a protein or to another RNA molecule within the cell such as pre-mRNA or tRNA or rRNA, or hybridises to, and interferes with the expression from a gene.

Conveniently, a gene expressing an antisense RNA may be constructed by inserting a coding sequence encoding a protein adjacent a promoter in the appropriate orientation such that the RNA complementary to mRNA. Suitably, the antisense RNA blocks expression of undesirable polypeptides such as oncogenes, for example *ras*, *bcl*, *src* or tumour suppressor genes such as *p53* and *Rb*.

30 It will be appreciated that it may be sufficient to reduce expression of the

undesirable polypeptide rather than abolish the expression.

It will be further appreciated that DNA sequences suitable for expressing as antisense RNA may be readily derived from publicly accessible 5 databases such as GenBank and EMBL.

In another embodiment of the invention, the gene replaces the function of a defective gene in the target cell.

10 There are several thousand inherited genetic diseases of mammals, including humans, that are caused by defective genes. Examples of such genetic diseases include cystic fibrosis, where there is known to be a mutation in the CFTR gene; Duchenne muscular dystrophy, where there is known to be a mutation in the dystrophin gene; sickle cell disease, 15 where there is known to be a mutation in the HbA gene. Many types of cancer are caused by defective genes, especially protooncogenes, and tumour-suppressor genes that have undergone mutation.

Thus, it is preferred that the virus or virus-like particle of the invention, 20 which may be useful in the treatment of cystic fibrosis, contains a functional CFTR gene to replace the function of the defective CFTR gene. Similarly, it is preferred that the virus or virus-like particle of the invention, which may be useful in the treatment of cancer, contains a functional protooncogene, or tumour-suppressor gene to replace the 25 function of the defective protooncogene or tumour-suppressor gene.

Examples of protooncogenes are *ras*, *src*, *bcl* and so on; examples of tumour-suppressor genes are *p53* and *Rb*.

30 By "gene" we mean a nucleic acid coding sequence that may contain

introns, or fragment thereof, or cDNA, or fragment thereof.

It will be appreciated that the gene will be introduced into a convenient place within the genome of the virus or virus-like particle and will contain  
5 a promoter and/or enhancer element to drive its expression.

It is preferred if the promoter and/or enhancer is selective for the cells to be targeted. Some examples of tissue or tumour specific promoters are given below but new ones are being discovered all of the time which will  
10 be useful in this embodiment of the invention.

The tyrosinase and TRP-1 genes both encode proteins which play key roles in the synthesis of the pigment melanin, a specific product of melanocytic cells. The 5' ends of the tyrosinase and tyrosinase-related  
15 protein (TRP-1) genes confer tissue specificity of expression on genes cloned downstream of these promoter elements.

The 5' sequences of these genes are described in Bradl, M. *et al* (1991) *Proc. Natl. Acad. Sci. USA* **88**, 164-168 and Jackson, I.J. *et al* (1991)  
20 *Nucleic Acids Res.* **19**, 3799-3804.

Prostate-specific antigen (PSA) is one of the major protein constituents of the human prostate secretion. It has become a useful marker for the detection and monitoring of prostate cancer. The gene encoding PSA and  
25 its promoter region which directs the prostate-specific expression of PSA have been described (Lundwall (1989) *Biochem. Biophys. Res. Comm.* **161**, 1151-1159; Riegman *et al* (1989) *Biochem. Biophys. Res. Comm.* **159**, 95-102; Brawer (1991) *Acta Oncol.* **30**, 161-168).

30 Carcinoembryonic antigen (CEA) is a widely used tumour marker,

especially in the surveillance of colonic cancer patients. Although CEA is also present in some normal tissues, it is apparently expressed at higher levels in tumorous tissues than in corresponding normal tissues. The complete gene encoding CEA has been cloned and its promoter region 5 analysed. A CEA gene promoter construct, containing approximately 400 nucleotides upstream from the translational start, showed nine times higher activity in the adenocarcinoma cell line SW303, compared with the HeLa cell line. This indicates that *cis*-acting sequences which convey cell type specific expression are contained within this region (Schrewe *et al* (1990) 10 *Mol. Cell. Biol.* **10**, 2738-2748).

The *c-erbB-2* gene and promoter have been characterised previously and the gene product has been shown to be over-expressed in tumour cell lines (Kraus *et al* (1987) *EMBO J.* **6**, 605-610).

15 The mucin gene, MUC1, contains 5' flanking sequences which are able to direct expression selectively in breast and pancreatic cell lines, but not in non-epithelial cell lines as taught in WO 91/09867.

20 The binding moiety allowing the virus or virus-like particle to bind to a target cell may be a polypeptide or oligosaccharide or lipid or any other molecule capable of binding specifically to the target cell.

It is preferred that the binding moiety is a polypeptide.

25 The molecule on the surface of the virus or virus-like particle to which the binding moiety is joined may be a polypeptide, oligosaccharide or lipid or any other molecule in the virus or virus-like particle coat. It is preferred that the molecule is a polypeptide.

If the binding moiety and the molecule on the surface of the virus or virus-like particle are both polypeptides then they may be linked together by any of the conventional ways of cross-linking polypeptides, such as those generally described in O'Sullivan *et al Anal. Biochem.* (1979) 100, 5 100-108. For example, the binding moiety may be enriched with thiol groups and the molecule on the surface of the virus or virus-like particle reacted with a bifunctional agent capable of reacting with those thiol groups, for example the N-hydroxysuccinimide ester of iodoacetic acid (NHIA) or N-succinimidyl-3-(2-pyridyldithio)propionate (SPDP). Amide 10 and thioether bonds, for example achieved with m-maleimidobenzoyl-N-hydroxysuccinimide ester, are generally more stable *in vivo* than disulphide bonds.

Other chemical procedures may be useful in joining oligosaccharide and 15 lipids to other oligosaccharides, lipids or polypeptides.

It is preferred that the binding moiety and the molecule on the surface of the virus or virus-like particle are both polypeptides that may be produced as a fusion by the techniques of genetic engineering. The use of genetic 20 engineering allows for the precise control over the fusion of such polypeptides.

Thus a further embodiment of the invention is a nucleotide sequence encoding the fusion of the binding moiety and the protein on the surface 25 of the virus or virus-like particle.

The nucleotide sequence encoding the fusion of the binding moiety and the protein on the surface of the virus or virus-like particle is preferably made by an alteration of the viral genome.

The nucleotide sequence may be synthesised *de novo* using solid phase phosphoramidite chemistry, but it is more usual for the nucleotide sequence to be constructed from two parts, the first encoding the binding moiety and the second the protein on the surface of the virus or virus-like

5 particle. The two parts may be derived from their respective genes by restriction endonuclease digestion or by other methods known by those skilled in the art such as the polymerase chain reaction.

A variety of methods have been developed to operatively link two

10 nucleotide sequences via complementary cohesive termini. For instance, synthetic linkers containing one or more restriction sites provide a method of joining the two DNA segment together. Each DNA segment, generated by endonuclease restriction digestion, is treated with bacteriophage T4 DNA polymerase of *E. coli* DNA polymerase I, enzymes that remove

15 protruding, 3'-single-stranded termini with their 3'-5'-exonuclease activities, and fill in recessed 3'-ends with their polymerizing activities.

The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large

20 molar excess of linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme

25 and lifted to an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the DNA segment.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including International

30 Biotechnologies Inc, New Haven, CN, USA.

A desirable way to generate the DNA encoding the fusion polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki *et al* (1988) *Science* 239, 487-491.

- 5 In this method each of the DNA molecules encoding the two polypeptides to be fused are enzymatically amplified using two specific oligonucleotide primers which themselves become incorporated into the amplified DNA. The said specific primers may contain restriction endonuclease recognition sites which may then be used to join the said two DNA molecules using
- 10 T4 DNA ligase as disclosed.

A particular feature of one aspect of the present invention is the modification of the virus or virus-like particle of the invention so that it no longer binds its host cell and so that it binds the target cell by virtue

15 of its binding moiety.

The host-cell receptor of adenovirus may be the penton fibre and that of influenza virus may be the haemagglutinin receptor.

- 20 These receptors may be modified by the insertion or deletion or substitution of amino acid residues that disrupt their host-cell binding function. It is preferred that the binding moiety for the target cell is joined to the host-cell receptor in such a manner that the binding moiety is capable of binding the target cell, the host-cell receptor is unable to
- 25 bind to the host cell and therefore the binding specificity of the virus or virus-like particle is modified. A further preference is that the portion of the host-cell receptor that is exposed on the surface of the virus or virus-like particle is replaced by the binding moiety, and that the portion of the host-cell receptor which promotes the uptake of viral DNA by the target
- 30 cell is retained. Suitably, the binding moiety is joined directly or

indirectly to the host-cell receptor by a spacer group.

Examples of spacer groups are polypeptide sequences of between 4 and 1000 amino acid residues.

5

Thus, in one embodiment of the invention the gene encoding the penton fibre in adenovirus is modified in such a way that the DNA encoding the surface-exposed portion is replaced by a DNA fragment encoding a ScFv, the ScFv being derived from an antibody which binds to a target cell 10 surface antigen.

Potential fusion sites within the penton fibre have been identified.

15 The adenovirus fibre is a trimer composed of three protomers. The amino terminal end (40 amino acids or so) of each participates in the formation of a tail that is closely associated with the penton (as opposed to the hexon) subunit of the capsid. High amino acid conservation is maintained between the different characterised serotypes.

20 Middle portions of each protomer form the shaft of the protein. This shaft is of variable length, depending upon serotype, and is composed of repeating units of 15 amino acids (for examples, serotypes have been identified with 6, 15 and 21 repeat units). These repeating units are not duplicates: rather than strict conservation of amino acid structure, there 25 is a general conservation of relative hydrophobicity. Some serotypes, for example, 40 and 41, have shafts composed of different length fibre proteins. This suggests a certain flexibility in structural constraints.

30 The carboxy-terminal ends (some 200 amino acids) associate to form a knob that is held erect a great distance (in molecular terms) from the

capsid.

Whilst the cellular receptor(s) and mechanisms of docking have not been firmly identified and elucidated, we propose that the most likely candidate 5 structure for cell binding is the knob. Thus, in one embodiment the whole knob of the penton fibre has been replaced with single chain antibody (ScFv) domains. The triplex structure implies that each fibre will thus end in three ScFvs. Additionally, the ScFv regions can be replaced with CDRs, or by non-antibody derived peptides, of known specificity or other 10 molecules that are capable of interacting specifically with the target cell.

Suitable fusion sites are therefore at the native junction between shaft and knob domains, or (should the DNA sequence prove to be more amenable) at any junction between repetitive units of the shaft. Preferably, the 15 minimum shaft length is not reduced beyond the smallest size naturally identified. There are thus at least 15 potential sites at which fusion could be contemplated.

Although it is preferred that the binding moiety forms the end of the fibre 20 thereby replacing the knob, the binding moiety may also be fused within the penton fibre sequence but still display its binding surfaces and bind to the target cell.

Suitably, the binding moiety may be fused to the knob and extend 25 externally to the knob structure.

In a further embodiment influenza virus haemmagglutinin is modified to incorporate a binding moiety. Influenza virus has seven or eight (depending on serotype) genetic segments, all negative strand RNA. 30 Suitably, a cDNA from the whole segment encoding haemmagglutinin is

constructed and modified by adding a promoter firing *backwards* across this segment so that negative strand RNA is made. Genetic fusions with a suitable binding molecule, as disclosed above, are made using standard recombinant DNA methods and a suitable cell line is stably transfected 5 with this gene construct. Infection of this transfected cell line with influenza virus and selection of reassorted genomes containing the new haemmagglutinin by infection of a normally resistant cell line that expresses a marker that can only be recognised by the new haemmagglutinin yields the desired virus comprising modified cell-binding 10 specificity.

A further aspect of the invention provides a method of producing in cell culture a virus or virus-like particle and then joining the binding moiety, as defined above, to the virus or virus-like particle.

15 A further aspect of the invention provides a method of producing in cell culture a virus or virus-like particle which has been genetically modified to express a binding moiety on its surface. The virus or virus-like particle is grown in its host prior to modification, but once the modification that 20 alters the binding specificity is made, the virus or virus-like particle is grown in the target cell. Thus, for example in the case where the binding moiety recognises a breast tumour cell antigen, the virus or virus-like particle is grown in breast tumour cell culture.

25 The virus or virus-like particles of the invention are administered in any suitable way, usually parenterally, for example intravenously, intraperitoneally or intravesically, in standard sterile, non-pyrogenic formulations of diluents and carriers, for example isotonic saline (when administered intravenously).

A further aspect of the invention provides a method of delivery of the virus or virus-like particle which contains a gene encoding a molecule having an indirectly cytotoxic function.

5 Suitably, the indirectly cytotoxic function is an enzyme that converts a prodrug to a toxic drug. With such a virus or virus-like particle, once the virus or virus-like particle has bound to the target cells, delivered its nucleic acid to the cells, and expressed the indirectly cytotoxic functions, which typically takes a day or so, the pro-drug is administered. The  
10 timing between administration of the virus or virus-like particle and the pro-drug may be optimised in a non-inventive way.

The dosage of the pro-drug will be chosen by the physician according to the usual criteria. The dosage of the virus or virus-like particle will  
15 similarly be chosen according to normal criteria, and in the case of tumour treatment, particularly with reference to the type, stage and location of tumour and the weight of the patient. The duration of treatment will depend in part upon the rapidity and extent of any immune reaction to the virus or virus-like particle.

20 Some of the viruses or virus-like particles either in themselves, or together with an appropriate pro-drug, are in principle suitable for the destruction of cells in any tumour or other defined class of cells selectively exhibiting a recognisable (surface) entity. Examples of types of cancer that may be  
25 treated using the viruses or virus-like particles are cancer of the breast, prostate, colon, rectum, ovary, testicle and brain. The compounds are principally intended for human use but could be used for treating other mammals including dogs, cats, cattle, horses, pigs and sheep.

30 The invention will now be described in detail with reference to the

following Figures and Examples in which:

5 Figure 1 shows (a) an unmodified (i.e. "naive") virus or virus-like particle able to bind to and infect its host cell but not a non-host cell, such as a target cell; and (b) a virus or virus-like particle with a modified binding specificity does not bind and infect its host cell but binds and infects a target cell; and (c) a virus or virus-like particle as in (b) modified further to contain a gene for gene therapy or cancer treatment.

10 Figure 2 shows (a) unmodified (naive) adenovirus; (b) adenovirus modified so that its penton fibres, which recognise the host cell, are replaced in part by antibody fragments which recognise the target cell; and (c) adenovirus as in (b) with further genetic material added to the viral DNA for gene therapy of cancer.

15 Figure 3 shows (a) influenza virus and (b) genetically-modified influenza virus wherein at least part of the haemagglutinin binding site is replaced by an antibody with anti-cancer cell binding activity.

20 Figure 4 shows (a) a retrovirus virus; and (b) as in (a) except the retrovirus has been modified further to express on its surface an anticancer cell-binding antibody fragment or an anticancer cell-binding peptide.

25 Figure 5 is a diagrammatic representation of a penton fibre indicating potential fusion sites within the fibre.

Figure 6 shows fusions between the DNA encoding the Ad5 fibre and an ScFv.

30 Figure 7 shows sequences of oligonucleotides used for amplifying the

ScFv. All oligonucleotides are presented 5' to 3', the reverse complement of FOR primers are shown and derived amino acid sequences are shown where relevant.

5 Figure 8 shows the construction of plasmid pRAS117.

Figure 9 shows the nucleotide and derived amino acid sequence between the *Hind*III and *Eco*RI sites of pRAS117.

10 Figure 10 shows a map of plasmid pRAS117.

Figure 11 is a diagrammatic representation of the construction of plasmid pRAS118.

15 Figure 12 shows the sequences of oligonucleotides for amplifying Ad5 fibre DNA fragments. All oligonucleotides are presented 5' → 3'. The reverse complements of FOR primers are shown. Derived amino acid sequences are shown where relevant.

20 Figure 13 shows the nucleotide sequence and deduced amino acid sequence between the *Hind*III site and *Eco*RI site of pRAS111.

Figure 14 gives a diagrammatic representation of constructing adenovirus carrying a cytotoxic gene.

25

Figure 15 gives the nucleotide and amino acid sequences of mouse and humanised HMFG1 variable regions.

Example 1: Fusion sites within the adenovirus Ad5 fibre for binding  
moieties including single chain Fv (ScFv)

The Ad5 DNA sequence co-ordinates used here are taken from:

5           ADRCOMPGE\_1: residues 1 to 32760  
and        ADRCOMPGE\_2: residues 32761-35935

These can be accessed by using program SEQ on the Intelligenetics database.

10

The sequence of Ad5 fibre can also be found in Chroboczek, J. and Jacrot, B. (1987) "The sequence of adenovirus fiber: Similarities and differences between serotypes 2 and 5" *Virology* 161, 549-554 and is available from the EMBL Database, Heidelberg, Germany under accession 15 name ADEFIB.

Fusion sequences between the shaft and the ScFv are shown in Fig. 6. The fusion sites are at the junctions of the repetitive units of the shaft. Shaft sequences are shown in normal typescript; ScFv sequences are 20 shown in italics. The DNA sequence between the *Pst*I and *Xba*I sites is unique to the ScFv used.

Fusion A is at the end of the first repetitive unit of the shaft (co-ordinates 31218-9), fusion B at the end of the second (31266-7), fusion C at the 25 third (31323-4), fusion D at the fourth (31368-9), fusion E at the fifth (31413-4), fusion F at the sixth (31458-9), fusion G at the seventh (31503-4), fusion H at the eighth (31551-2), fusion I at the ninth (31596-7), fusion J at the tenth (31641-2), fusion K at the eleventh (31692-3), fusion L at the twelfth (31737-8), fusion M at the thirteenth (31787-8), fusion N 30 at the fourteenth (31836-7), fusion O at the fifteenth (31884-5), fusion P

at the sixteenth (31929-30), fusion Q at the seventeenth (31995-6), fusion R at the eighteenth (32040-1), fusion S at the nineteenth (32103-4), fusion T at the twentieth (32151-2), fusion U at the twenty-first (32199-200), and fusion V is at the end of the twenty-second repetitive unit of the shaft  
5 (32244-5), the junction between shaft and knob.

**Example 2: Preparation of adenovirus expressing an ScFv on its surface**

10 The genetically modified fibre is introduced into the Ad5 genome by: (a) replacing the fibre gene of plasmid pE4 with the modified fibre by standard recombinant DNA technology and (b) reconstituting the virus by recombination.

15 pE4 is a plasmid containing the right hand half of the Ad5 genome, and which has served as the source of the Ad5 fibre gene that we have used. It was provided by Dr Keith Leppard, Biological Sciences, University of Warwick, Coventry, CV4 7AL who has supplied details of its structure. If it is introduced into mammalian cells that contain the remainder of the  
20 Ad5 genome, then it is possible to obtain recombinants containing the modification. Most human cell lines can be used for the recombination but HeLa cells are preferred.

The plasmid pE4 is readily made in the following way. A derivative of  
25 pBR322 is made by digesting with *Bst*N1 and rejoining using *Xba*I linkers such that the *Bst*N1 fragment corresponding to positions 1442-2502 in the pBR322 sequence is removed. DNA from the adenovirus Ad5 strain 309 described by Jones & Shenk (1979) *Cell* 17, 683-689 is isolated and deproteinated. This DNA is then ligated to *Cla*I linkers and cut with  
30 *Eco*RI and *Cla*I. The *Cla*I-*Eco*RI fragment corresponding to the region

of 76% of the Ad5 genome to the right hand end is isolated and cloned into the *Eco*RI-*Cl*ai sites of the above-mentioned pBR322 derivative to form pE4.

5 Adenovirus Type 5 and HeLa cells are available from the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852-1776, USA under accession numbers ATCC VR-5 and ATCC CCL-2.

#### Construction of plasmid pRAS117

10

Oligonucleotide primers LEADHBACK and LEAdbFOR (Figure 7) were used for PCR-mediated amplification of the DNA segment extending from the *Hind*III site of plasmid pRAS111, over the Shine-Dalgarno sequence and the *pe*B leader sequence to the *Pst*I site in the ScFv. LEAdbFOR 15 directs the incorporation of a *Bgl*II site immediately after the *pe*B leader sequence. DNA (100 ng) from plasmid pRAS111 was subjected to 24 rounds of amplification, (94°C, 1 min; 65°C, 1.5 min and 72°C, 2 min) in a 50 µl reaction volume containing 25 pmol of each primer, 250 mM of each dNTP, 67 mM Tris-HCl (pH 8.8), 17 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.5 mM 20 MgCl<sub>2</sub>, 200 mg.ml<sup>-1</sup> gelatine and 5 units of *Thermus aquaticus* (Taq) polymerase (Cetus) overlaid with 25 µl paraffin oil. After the reaction, oil was removed by extraction with 500 µl chloroform. The sample was loaded on a 2% agarose gel, and the amplified fragment was electrophoresed on to a piece of NA45 paper (Schleicher and Schuell). 25 Bound DNA was subsequently eluted by immersion in 400 µl 1M NaCl made in TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) for 30 min at 70°C. To this was added 800 µl ethanol, and after incubation (2 h, -20°C) the DNA was collected by centrifugation. The pellet was taken up in 50 µl T/E.

30

One fifth (10  $\mu$ l) of the purified amplified fragment was cut with the restriction enzymes *Hind*III and *Pst*I, in a total volume of 20  $\mu$ l 50 mM Tris-HCl, pH 7.5, 10 MgCl<sub>2</sub>, 100 mM NaCl, 1 mM dithioerythreitol containing 10 units of each enzyme. After incubation (1 h, 37°C) the 5 reaction was stopped by incubation at 70°C for 15 minutes.

The trimmed amplified fragment was cloned between the *Hind*III and *Pst*I sites of pUC8, to generate plasmid pRAS117.

10 Plasmid pUC8 (1  $\mu$ g) was cut with *Hind*III and *Pst*I, in a total volume of 20  $\mu$ l 50 mM Tris-HCl, pH 7.5, 10 MgCl<sub>2</sub>, 100 mM NaCl, 1 mM dithioerythreitol containing 10 units of each enzyme. After incubation (1 h, 37°C) the reaction was stopped by incubation at 70°C for 15 minutes.

15 The ligation reaction contained 1.5  $\mu$ l of pUC8/*Hind*III, *Pst*I and 3  $\mu$ l of the amplified leader/*Hind*III, *Pst*I in a total volume of 15  $\mu$ l containing 70 mM Tris-HCl pH 7.5, 7 mM MgCl<sub>2</sub>, 0.7 mM rATP, 4 mM dithiothreitol, 0.5 mg.ml<sup>-1</sup> BSA and 10 units of T4 DNA ligase. After incubation (2 h, at room temperature), the reaction was stopped by the addition of 1  $\mu$ l 500 20 mM EDTA, pH 8.0 and 14  $\mu$ l H<sub>2</sub>O.

This ligation mix was used to transform *E. coli*.

An aliquot (5  $\mu$ l) of this ligation mix was used to transform a 200  $\mu$ l 25 aliquot of commercially available competent *E. coli* K12 DH58, *l*<sub>α</sub>*F* (Life Sciences Inc). After incubation (30 min, 0°C), heat shock (2 min, 42°C), addition of 800  $\mu$ l L-broth and recovery (37°C, 1 h), cells (100  $\mu$ l) were spread on L-agar plates containing 100  $\mu$ g.ml<sup>-1</sup> ampicillin containing 50 mM IPTG (isopropyl- $\beta$ -D-galactopyranoside) and 100  $\mu$ g.ml<sup>-1</sup> X-Gal (5-30 bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside). Cells were grown

overnight at 37°C, and individual colonies were transferred to fresh L-agar/ampicillin plates. After 6 h growth, colonies were used to inoculate 5 ml aliquots of L-broth containing 100  $\mu\text{g.ml}^{-1}$  ampicillin. These cells were grown overnight with shaking at 37°C, and used as a source of 5 plasmid DNA.

These cells were used as a source of plasmid DNA.

Harvested cells were suspended in 360  $\mu\text{l}$  of SET (50 mM sucrose, 10 10 mM EDTA, 100 mM Tris-HCl, pH 7.5) containing 2  $\text{mg.ml}^{-1}$  hen egg lysozyme, transferred to a 1.5 ml microfuge tube, and diluted by addition of 300  $\mu\text{l}$  10% Triton X-100. After floating on boiling water for 2 min and cooling for a further minute in ice/water, denatured cell debris was removed by centrifugation (14,000  $\times g$ , 20 min) in a microcentrifuge. 15 The majority of the soluble remaining proteins were removed by addition of 300  $\mu\text{l}$  7.5 M ammonium acetate and centrifugation (14,000  $\times g$ , 10 min). Nucleic acids were precipitated by addition of 720  $\mu\text{l}$  cold (-20°C) isopropanol and centrifugation (14,000  $\times g$ , 10 min). After rinsing the 20 pellets with ethanol and drying, DNA was solubilised in 60  $\mu\text{l}$  TE containing 170  $\mu\text{g.ml}^{-1}$  RNase A.

Restriction enzyme digestions on 5  $\mu\text{l}$  aliquots, using the enzymes *Hind*III and *Bg*II identified which of these plasmids were pRAS117. The construction scheme is shown in Fig. 8. The nucleotide and derived 25 amino acid sequences between the *Hind*III and *Eco*RI sites of pRAS117 are shown in Fig. 9. A map of plasmid pRAS117 is provided in Fig. 10.

The nucleotide sequence of the relevant portion of pRAS111, between the *Hind*III site and *Eco*RI, site is given in Figure 13.

**Construction of plasmid pRAS118 (Figure 11)**

The 130bp *Hind*III-*Pst*I fragment of pRAS117 was used to replace the corresponding fragment of pRAS111, to generate plasmid pRAS118. An 5 aliquot (2 µg) of pRAS111 DNA was cut with *Hind*III and *Pst*I in the conditions used previously, the large fragment was isolated by electrophoresis onto NA45 paper, as described previously, and the DNA was suspended in 10 µl of TE. An aliquot (10 µl) of pRAS117 DNA was cut with *Hind*III and *Pst*I in the conditions used previously, and the small 10 fragment was isolated by electrophoresis onto NA45 paper, as described previously, and the DNA was suspended in 10 µl of TE.

The isolated pRAS111/*Hind*III/*Pst*I large fragment (1.5 µl) and the isolated pRAS117/*Hind*III/*Pst*I small fragment (3 µl) were mixed and ligated in the 15 conditions previously described.

Transformations, colony handling and DNA preparations were as previously described.

20 Restriction enzyme digestions on 5 µl aliquots, using the enzymes *Hind*III, *Pst*I and *Bg*II identified which of these plasmids were pRAS118. This encodes a NIP-reactive ScFv with a *Bg*II cloning site immediately downstream of the *pefB* leader, suitable for inserting fragments of DNA from Ad5 fibre (and also suitable for fusion of any other desired fusion 25 functions).

**Amplification of Ad5 fibre DNA fragments**

Fragments of DNA from Ad5 fibre were amplified by PCR using 30 oligonucleotide TAILbBACK and oligonucleotide FIBREPFOR,

FIBRE3FOR, FIBRE6FOR, FIBRE9FOR, FIBRE12FOR, FIBRE15FOR, FIBRE18FOR, FIBRE21FOR or FIBRE22FOR. Oligonucleotide sequences can be found in Fig. 12.

5 TAILdBACK directs the incorporation of a *Bg*II site at the base of the fibre, and the FIBREnFOR series primers direct the incorporation of a *Pst*I site at the junctions of repetitive shaft units 3-4 (FIBRE3FOR), 6-7 (FIBRE6FOR), 9-10 (FIBRE9FOR), 12-13 (FIBRE12FOR), 15-16 (FIBRE15FOR), 18-19 (FIBRE18FOR), 21-22 (FIBRE21FOR), between 10 unit 22 and the knob (FIBRE22FOR) or at the end of the knob sequence (FIBREPFOR).

#### Fusion of fibre and ScFv

15 The amplified segments of fibre are trimmed with *Bg*II and *Pst*I and ligated between the *Bg*II and *Pst*I sites of plasmid pRAS118. This gives a range of fusions under the transcriptional control of the T7 promoter. Colonies are recovered after transformation of a suitable *E. coli* strain, such as DH5, which does not permit expression of the fusions.

20

#### Screening

Colonies containing candidates for fusion are identified by restriction digestion of their plasmid DNAs. These candidate DNAs are used to 25 transform a suitable *E. coli* strain, such as BL21 (DE3), that contains a chromosomal insertion of T7 polymerase under *lac* control. In these cells, induction of expression of T7 polymerase using the gratuitous inducer IPTG causes expression of the fusion proteins. Soluble NIP-reactive material is identified in colonies with correctly assembled fusions. The 30 DNA of these is identified and the NIP-reactive ScFv derived from

pRAS111 are replaced with a cell-binding ScFv.

**Replacing the fibre:ScFv in plasmid pE4**

5 There is a *Hind*III site approximately half-way along the fibre gene. Fusions with long fibres also contain this *Hind*III site. The fusion is introduced at this site.

10 **Recombination *in vivo* of plasmid pE4-ScFv with the adenovirus genome**

To obtain virus particles expressing the ScFv on the penton fibre suitable cells, such as 293 cells, are cotransfected with plasmid pE4-ScFv and plasmid pFG173 as described in Mittal *et al* (1993) *Virus Res.* 28, 67-90, 15 incorporated herein by reference. Since neither pFG173 nor pE4-ScFv individually is able to generate virus progeny, on transfection of 293 cells viable virus progeny are only produced by *in vivo* recombination between these two plasmids resulting in rescue of the penton fibre-ScFv fusion into the Ad5 genome.

20 293 cells are human transformed primary embryonal cells available from the ATCC under accession number ATCC CRL 1573.

25 The adenovirus particles made in this way express a NIP-binding ScFv on their surface. Such particles are useful in a two-step targeting approach wherein a target-cell specific binding moiety, such as those identified in Tables 1 and 2, are joined to NIP molecule and targeted to a cell. Once they have localized to the target cell within the patient, the adenovirus displaying NIP-binding ScFv is administered to the patient and binds to the 30 NIP.

Example 3: Insertion of a cytotoxic gene into the E3 region of adenovirus Ad5

In preparation for rescue of the cytotoxic gene into the E3 region of Ad5,  
5 the cytotoxic coding sequences were first inserted into a cassette containing the SV40 early promoter and poly A addition sequences to give plasmid pTOX as shown in Figure 14.

10 To obtain virus with the cytotoxic gene and SV40 regulatory sequences in the E3 region, 293 cells are cotransfected with plasmids pTOX and pFG173 (Fig 14). The plasmid pFG173 is constructed from pFG140, an infectious plasmid containing the Ad5 d1309 genome in circular form by inserting a kan' gene at the *Eco*RI site as 75.9 m.u. as described in Graham (1984) *EMBO J.* 3, 2917-2922 and Mitall *et al* (1993) *Virus Res.* 28, 67-15 90.

20 Since neither pFG173 nor pTOX individually is able to generate infectious virus progeny, on transfection of 293 cells viable virus progeny are only produced by *in vivo* recombination between these two plasmids resulting in rescue of the E3 insert into the Ad5 genome.

Viral plaques obtained after cotransfection are isolated and expanded in 293 cells and viral DNA was analyzed on an agarose gel after digestion with *Hind*III. The structure of the desired Ad5-cytotoxic gene recombinant is verified by the presence of diagnostic fragments. One recombinant is plaque purified and used for further study.

**Legend to Figure 14**

30 The plasmid pFG173 contains the entire Ad5 genome, except for a 3.2 kb

sequence spontaneously deleted between m.u. 75.9-84.9. Plasmids pTOX and pFG173 were used for cotransfection of 293 cells to rescue, by *in vivo* recombination, the cytotoxic gene flanked by SV40 regulatory sequences in the E3 region of Ad5. The resulting Ad5-cytotoxic gene recombinant 5 was named Ad5-TOX. The relative positions of *Hind*III and *Xba*I restriction sites of the Ad5-TOX genome are shown. The position and orientation of the SV40 promoter, the cytotoxic gene, and the SV40 polyadenylation signal are shown below. Solid bars: luciferase gene; open bars: SV40 promoter and SV40 polyadenylation signal; hatched 10 bars: amp' and kan' genes.

The cytotoxic gene is the cDNA for thymidine kinase.

Other cytotoxic genes are inserted into the E3 region of Ad5 in an 15 analogous manner.

**Example 4: Single chain Fv from the mouse monoclonal antibody HMFG1 and humanised monoclonal antibody Hu HMFG1**

20 The nucleotide sequences encoding the V<sub>H</sub> heavy chains and V<sub>L</sub> light chains of HMFG1 and Hu HMFG1 are shown in Figure 15 and are given in Verhoeven *et al* (1993) *Immunology* 78, 364-370, incorporated herein by reference.

25 **Legend to Figure 15**

Nucleotide and amino acid sequences of mouse and reshaped HMFG1 variable regions. (a) Heavy chain variable region sequences for mouse and reshaped HMFG1 (Mo V<sub>H</sub>-HMFG1 and Hu V<sub>H</sub>-HMFG1); (b) mouse 30 and reshaped light chain variable regions respectively (Mo V<sub>L</sub>-HMFG1 and

Hu V<sub>H</sub>-HMFG1). Amino acids numbering and definition of the CDR and framework regions are from Kabat *et al* (1987) *Sequences of Proteins of Immunological Interest*, Edn 4, US Dept of Health and Human Services Public Health Service, NIH, Bethesda, MD 20892, USA.

5

The methods described by Bird *et al* (1988) *Science* 242, 423 or Huston *et al* (1988) *Proc. Natl. Acad. Sci. USA* 85, 5879 are applied to the nucleotide sequences described in Figure 15 to generate genes encoding ScFv for HMFG1 and ScFv for Hu HMFG1. These genes are fused 10 individually into the adenovirus penton fibre gene as described in Examples 1 and 2.

The amino acid sequences of the V<sub>H</sub> and V<sub>L</sub> chains of H17E2 are disclosed in "Monoclonal antibodies - applications in clinical oncology", pages 37-15 43, 1991, A.A. Epenetos, ed., Chapman & Hall, UK.

Nucleotide sequences encoding the V<sub>H</sub> and V<sub>L</sub> chains are readily derived from the amino acid sequence using the genetic code and an ScFv can be made from the sequences using the methods of Bird *et al* or Huston *et al* 20 as described above.

Key to Sequence Listing

Name	SEQ ID No.	
	Nucleotide Sequence	Polypeptide Sequence

	Fusion A	1	2
5	Fusion B	3	4
	Fusion C	5	6
	Fusion D	7	8
	Fusion E	9	10
	Fusion F	11	12
	Fusion G	13	14
	Fusion H	15	16
10	Fusion I	17	18
	Fusion J	19	20
	Fusion K	21	22
	Fusion L	23	24
	Fusion M	25	26
15	Fusion N	27	28
	Fusion O	29	30
	Fusion P	31	32
	Fusion Q	33	34
	Fusion R	35	36
20	Fusion S	37	38
	Fusion T	39	40
	Fusion U	41	42
	Fusion V	43	44
	<i>Xho-Eco</i>	45	-
25	LEADHBACK	46	-
	LEADbFOR	47	48
	pRAS117	49	50
	TAILbBACK	51	52
	FIBRE3FOR	53	54
30	FIBRE6FOR	55	56
	FIBRE9FOR	57	58
	FIBRE12FOR	59	60
	FIBRE15FOR	61	62
	FIBRE18FOR	63	64
35	FIBRE21FOR	65	66
	FIBRE22FOR	67	68
	FIBREPFOR	69	70
	pRAS111	71	72
	MoV <sub>H</sub>	73	74
40	MoV <sub>z</sub>	75	76
	HuV <sub>H</sub>	77	78
	HuV <sub>z</sub>	79	80

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: Imperial Cancer Research Technology Limited
- (B) STREET: Sardinia House, Sardinia Street
- (C) CITY: London
- (E) COUNTRY: United Kingdom
- (F) POSTAL CODE (ZIP): WC2A 3NL

## (ii) TITLE OF INVENTION: Compounds to target cells

## (iii) NUMBER OF SEQUENCES: 80

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iii) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus
- (B) STRAIN: Ad5

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

CCT CTA GTT ACC TCC AAT GTG CAG CTG CAG
30
Pro Leu Val Thr Ser Asn Val Gln Leu Gln
 1           5           10

```

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Pro Leu Val Thr Ser Asn Val Gln Leu Gln

```

1 5 10

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CTC TCT CTG GAC GAG GCC GTG CAG CTG CAG  
 30  
 Leu Ser Leu Asp Glu Ala Val Gln Leu Gln  
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Leu Ser Leu Asp Glu Ala Val Gln Leu Gln  
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCT CTC AAA AAA ACC AAG GTG CAG CTG CAG  
30  
Pro Leu Lys Lys Thr Lys Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Pro Leu Lys Lys Thr Lys Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CCC CTC ACA GTT ACC TCA GTG CAG CTG CAG  
30  
Pro Leu Thr Val Thr Ser Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Pro Leu Thr Val Thr Ser Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus
- (B) STRAIN: Ad5

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCT CTA ATG GTC GCG GGC GTG CAG CTG CAG  
30

Pro Leu Met Val Ala Gly Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Pro Leu Met Val Ala Gly Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus

(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCG CTA ACC GTG CAC GAC GTG CAG CTG CAG  
 30  
 Pro Leu Thr Val His Asp Val Gln Leu Gln  
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Pro Leu Thr Val His Asp Val Gln Leu Gln  
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Adenovirus

(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCC CTC ACA GTG TCA GAA GTG CAG CTG CAG  
 30  
 Pro Leu Thr Val Ser Glu Val Gln Leu Gln  
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Pro Leu Thr Val Ser Glu Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus
- (B) STRAIN: Ad5

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTC ACC ACC GAT AGC GTG CAG CTG CAG  
30  
Leu Thr Thr Thr Asp Ser Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Leu Thr Thr Thr Asp Ser Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

50

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CCT CTA ACT ACT GCC ACT GTG CAG CTG CAG  
30  
Pro Leu Thr Thr Ala Thr Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Pro Leu Thr Thr Ala Thr Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CCC ATT TAT ACA CAA RAT GTG CAG CTG CAG  
30  
Pro Ile Tyr Thr Gln Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Pro Ile Tyr Thr Gln Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAT GTA ACA GAC GAC CTA GTG CAG CTG CAG  
30  
His Val Thr Asp Asp Leu Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

His Val Thr Asp Asp Leu Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: N

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus
- (B) STRAIN: Ad5

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGT GTG ACT ATT AAT AAT GTG CAG CTG CAG  
30  
Gly Val Thr Ile Asn Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Gly Val Thr Ile Asn Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Adenovirus
- (B) STRAIN: Ad5

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGT TTT GAT TCA CAA CGC GTG CAG CTG CAG  
30  
Gly Phe Asp Ser Gln Gly Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Gly Phe Asp Ser Gln Gly Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AGG ATT GAT TCT CAA AAC GTG CAG CTG CAG  
30  
Arg Ile Asp Ser Gln Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Arg Ile Asp Ser Gln Asn Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..30
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TTT GAT GCT CAA AAC CAA GTG CAG CTG CAG  
30  
Phe Asp Ala Gln Asn Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Phe Asp Ala Gln Asn Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..30
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CTT TTT ATA AAC TCA GCC GTG CAG CTG CAG  
30  
Leu Phe Ile Asn Ser Ala Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Leu Phe Ile Asn Ser Ala Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TCA AAC AAT TCC AAA AAC GTG CAG CTG CAG  
30  
Ser Asn Asn Ser Lys Asn Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Ser Asn Asn Ser Lys Asn Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

56

(iii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GGG TTG ATG TTT GAC GCT GTG CAG CTG CAG

30

Gly Leu Met Phe Asp Ala Val Gln Leu Gln

1

5

10

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

Gly Leu Met Phe Asp Ala Val Gln Leu Gln

1

5

10

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CCT AAT GCA CCA AAC ACA GTG CAG CTG CAG

30

Pro Asn Ala Pro Asn Thr Val Gln Leu Gln

1

5

10

## (2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

Pro Asn Ala Pro Asn Thr Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Adenovirus  
(B) STRAIN: Ad5

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CTA GAA TTT GAT TCA AAC GTG CAG CTG CAG  
30  
Leu Glu Phe Asp Ser Asn Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

Leu Glu Phe Asp Ser Asn Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

CTT AGT TTT GAC AGC ACA GTG CAG CTG CAG  
 30  
 Leu Ser Phe Asp Ser Thr Val Gln Leu Gln  
 1   5   10

(2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

Leu Ser Phe Asp Ser Thr Val Gln Leu Gln  
 1   5   10

(2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iii) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATT GAT TAA CTA ACT TTG GTG CAG CTG CAG  
 30

Ile Asp Lys Leu Thr Leu Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Ile Asp Lys Leu Thr Leu Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iii) ANTI-SENSE: NO

- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Adenovirus
  - (B) STRAIN: Ad5

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CTCGAGTAAAT AGAAATTC  
18

## (2) INFORMATION FOR SEQ ID NO: 46:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iii) ANTI-SENSE: NO

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

AGCTTAAGCTT GCATGCAATT TC  
22

## (2) INFORMATION FOR SEQ ID N : 47:

- (i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CCA GCG ATG GCC AGA TCT CAG CTG CAG AGCT  
31  
Pro Ala Met Ala Arg Ser Gln Leu Gln  
1 5

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

Pro Ala Met Ala Arg Ser Gln Leu Gln  
1 5

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 40..132

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AAGCTTGAT GCAAATTCTA TTTCAGGAG ACAGTCATA ATG AAA TAC CTA TTG  
54  
Met Lys Tyr Leu Leu  
1 5

CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC GCT GCC CAA CCA GCG ATG  
 102  
 Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala Ala Gln Pro Ala Met  
 10 15 20  
 GCC AGA TCT CAG CTG CAG GTC GAC GGA TCC  
 132  
 Ala Arg Ser Gln Leu Gln Val Asp Gly Ser  
 25 30

## (2) INFORMATION FOR SEQ ID NO: 50:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala  
 1 5 10 15  
 Ala Gln Pro Ala Met Ala Arg Ser Gln Leu Gln Val Asp Gly Ser  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO: 51:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iii) ANTI-SENSE: NO

- (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 5..28

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

AGCT AGA TCT ATG AAG CCC GCA AGA CCG  
 28  
 Arg Ser Met Lys Arg Ala Arg Pro  
 1 5

## (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

Arg Ser Met Lys Arg Ala Arg Pro  
1 5

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CCT CTC AAA AAA ACC AAG CAG GTG CAG CTG CAG CAGCCTGG  
41  
Pro Leu Lys Lys Thr Lys Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

Pro Leu Lys Lys Thr Lys Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..34

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

C CCG CTA ACC GTG CAC GAC CAG GTG CAG CTG CAG CAGCCTGG

42

Pro Leu Thr Val His Asp Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Pro Leu Thr Val His Asp Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 57:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CCT CTA ACT ACT GCC ACT CAG GTG CAG CTG CAG CAGCCTGG

41

Pro Leu Thr Thr Ala Thr Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 58:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

Pro Leu Thr Thr Ala Thr Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs

64

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GGT GTG ACT ATT AAT AAT CAG GTG CAG CTG CAG GACCCCTGG  
41  
Gly Val Thr Ile Asn Asn Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

Gly Val Thr Ile Asn Asn Gln Val Gln Leu Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..36

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CCG TTT GAT GCT CAA AAC CAA CAG GTG CAG CTG CAG CAGCC  
41  
Pro Phe Asp Ala Gln Asn Gln Gln Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

Pro Phe Asp Ala Gln Asn Gln Gln Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

GGG TTG ATG TTT GAC GCT CAG GTG CAG CTG CAG CAGCC  
38  
Gly Leu Met Phe Asp Ala Gln Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

Gly Leu Met Phe Asp Ala Gln Val Gln Leu Gln  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 40 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

GC CTT AGT TTT GAC AGC ACA CAG GTG CAG CTG CAG CAGCC  
 40  
 Leu Ser Phe Asp Ser Thr Gln Val Gln Leu Gln  
 1                   5                           10

(2) INFORMATION FOR SEQ ID NO: 66:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

Leu Ser Phe Asp Ser Thr Gln Val Gln Leu Gln  
 1                   5                           10

(2) INFORMATION FOR SEQ ID NO: 67:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

GGA AAC AAA AAT AAT GAT AAG CTA ACT TTG CAG GTG CAG CTG CAG  
 45  
 Gly Asn Lys Asn Asn Asp Lys Leu Thr Leu Gln Val Gln Leu Gln  
 1                   5                           10                           15

CAGCC

50

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

Gly Asn Lys Asn Asn Asp Lys Leu Thr Leu Gln Val Gln Leu Gln  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 69:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iii) ANTI-SENSE: NO

- (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..17

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

CA TAC ATT GCC CAA GAA TAACAGGTGC AGCTGCCAGCA GCCTGG  
43  
Tyr Ile Ala Gln Glu  
1 5

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

Tyr Ile Ala Gln Glu  
1 5

(2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 858 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO

- (iii) ANTI-SENSE: NO

## (ix) FEATURE:

(A) NAME/KEY: CDS  
 (B) LOCATION: 40..846

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

AAGCTTGAT GCAAATTCTA TTTCAAGGAG ACAGTCATA ATG AAA TAC CTA TTG  
 54  
 Met Lys Tyr Leu Leu  
 1 5

CCT ACG GCA GCC GCT GGA TTG TTA TTA CTC GCT GCC CAA CCA GCG ATG  
 102  
 Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala Ala Gln Pro Ala Met  
 10 15 20

GCC CAG GTG CAG CTG CAG CAG CCT GGG GCT GAG CTT GTG AAG CCT GGG  
 150  
 Ala Gln Val Gln Leu Gln Gln Pro Gly Ala Glu Leu Val Lys Pro Gly  
 25 30 35

GCT TCA GTG CTG TCC TGC AAG GCT TCT GGC TAC ACC TTC ACC ACC  
 198  
 Ala Ser Val Lys Leu Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser  
 40 45 50

TAC TGG ATG CAC TGG GTG AAG CAG AGG CCT GGN CGA GGC CTT GAG TGG  
 246  
 Tyr Trp Met His Trp Val Lys Gln Arg Pro Gly Arg Gly Leu Glu Trp  
 55 60 65

ATT GGA AGG ATT GAT CCT AAT AGT GGT ACT AAG TAC AAT GAG AAG  
 294  
 Ile Gly Arg Ile Asp Pro Asn Ser Gly Gly Thr Lys Tyr Asn Glu Lys  
 70 75 80 85

TTC AAG AGC AAG GCC ACA CTG ACT GTA GAC AAA CCC TCC AGC ACA GCC  
 342  
 Phe Lys Ser Lys Ala Thr Leu Thr Val Asp Lys Pro Ser Ser Thr Ala  
 90 95 100

TAC ATG CAG CTC AGC AGC CTG ACA TCT GAG GAC TCT GCG GTC TAT TAT  
 390  
 Tyr Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr  
 105 110 115

TGT GCA AGA TAC GAT TAC TAC GGT AGT AGC TAC TTT GAC TAC TGG GGC  
 438  
 Cys Ala Arg Tyr Asp Tyr Tyr Gly Ser Ser Tyr Phe Asp Tyr Trp Gly  
 120 125 130

CAA GGG ACC ACG GTC ACC GTC TCC TCA GGT GGA GGC GGT TCA GCC GGA  
 486  
 Gln Gly Thr Thr Val Thr Val Ser Ser Gly Gly Gly Ser Gly Gly  
 135 140 145

GGT GGC TCT GGC GGT GGC GGA TCC CAG GCT GTT GTG ACT CAG GAA TCT  
 534  
 Gly Gly Ser Gly Gly Gly Ser Gln Ala Val Val Thr Gln Glu Ser  
 150 155 160

GCA CTC ACC ACA TCA CCT GGT GAA ACA GTC ACA CTC ACT TGT CGC TCA  
 582  
 Ala Leu Thr Thr Ser Pro Gly Glu Thr Val Thr Leu Thr Cys Arg Ser

69

170

175

180

AGT ACT GGG GCT GTT ACA ACT AGT AAC TAT GCC AAC TGG GTC CAA GAA  
 630  
 Ser Thr Gly Ala Val Thr Thr Ser Asn Tyr Ala Asn Trp Val Gln Glu  
 185 190 195  
 AAA CCA GAT CAT TTA TTC ACT GGT CTA ATA GGT GGT ACC AAC AAC CGA  
 678  
 Lys Pro Asp His Leu Phe Thr Gly Leu Ile Gly Gly Thr Asn Asn Arg  
 200 205 210  
 GCT CCA GGT GTT CCT GCC AGA TTC TCA GGC TCC CTG ATT GGA GAC AAG  
 726  
 Ala Pro Gly Val Pro Ala Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys  
 215 220 225  
 GCT GCC CTC ACC ATC ACA GGG GCA CAG ACT GAG GAT GAG GCA ATA TAT  
 774  
 Ala Ala Leu Thr Ile Thr Gly Ala Gln Thr Glu Asp Glu Ala Ile Tyr  
 230 235 240 245  
 TTC TGT GCT CTA TGG TAC AGC AAC CRC TGG GTG TTC GGT GGA GGA ACC  
 822  
 Phe Cys Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gly Thr  
 250 255 260  
 AAA CTG ACT GTC CTA GGT CTC GAG TAATAGAAT TC  
 858  
 Lys Leu Thr Val Leu Gly Leu Glu  
 265

## (2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 269 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala  
 1 5 10 15  
 Ala Gln Pro Ala Met Ala Gln Val Gln Leu Gln Gln Pro Gly Ala Glu  
 20 25 30  
 Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Lys Ala Ser Gly  
 35 40 45  
 Tyr Thr Phe Thr Ser Tyr Trp Met His Trp Val Lys Gln Arg Pro Gly  
 50 55 60  
 Arg Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Asn Ser Gly Gly Thr  
 65 70 75 80  
 Lys Tyr Asn Glu Lys Phe Lys Ser Lys Ala Thr Leu Thr Val Asp Lys  
 85 90 95  
 Pro Ser Ser Thr Ala Tyr Met Gln Leu Ser Ser Leu Thr Ser Glu Asp  
 100 105 110  
 Ser Ala Val Tyr Tyr Cys Ala Arg Tyr Asp Tyr Tyr Gly Ser Ser Tyr

70

115

120

125

Phe Asp Tyr Trp Gly Gln Gly Thr Thr Val Thr Val Ser Ser Gly Gly  
 130 135 140

Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gln Ala Val  
 145 150 155 160

Val Thr Gln Glu Ser Ala Leu Thr Thr Ser Pro Gly Glu Thr Val Thr  
 165 170 175

Leu Thr Cys Arg Ser Ser Thr Gly Ala Val Thr Thr Ser Asn Tyr Ala  
 180 185 190

Asn Trp Val Gln Glu Lys Pro Asp His Leu Phe Thr Gly Leu Ile Gly  
 195 200 205

Gly Thr Asn Asn Arg Ala Pro Gly Val Pro Ala Arg Phe Ser Gly Ser  
 210 215 220

Leu Ile Gly Asp Lys Ala Ala Leu Thr Ile Thr Gly Ala Gln Thr Glu  
 225 230 235 240

Asp Glu Ala Ile Tyr Phe Cys Ala Leu Trp Tyr Ser Asn His Trp Val  
 245 250 255

Phe Gly Gly Gly Thr Lys Leu Thr Val Leu Gly Leu Glu  
 260 265

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 354 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mouse

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..354

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CAG GTT CAG CTG CAG CAG TCT GGA GCT GAG CTG ATG AAG CCT GGG GCC  
 48  
 Gln Val Gln Leu Gln Gln Ser Gly Ala Glu Leu Met Lys Pro Gly Ala  
 1 5 10 15

TCA GTC AAG ATA TCC TGC AAG GCT ACT GCC TAC ACA TTC AGT GCC TAC  
 96  
 Ser Val Lys Ile Ser Cys Lys Ala Thr Gly Tyr Thr Phe Ser Ala Tyr  
 20 25 30

TGG ATA GAG TGG GTA AAG CAG AGG CCT GGA CAT GCC CTT GAG TGG ATT  
 144  
 Trp Ile Glu Trp Val Lys Gln Arg Pro Gly His Gly Leu Glu Trp Ile

35	40	45
GGA GAG ATT TTA CCT GGA AGT AAT AAT TCT AGA TAC AAT GAG AAG TTC		
192		
Gly Glu Ile Leu Pro Gly Ser Asn Asn Ser Arg Tyr Asn Glu Lys Phe		
50	55	60
AAG GGC AAG GCC ACA TTC ACT GCT GAT ACA TCC TCC AAC ACA GCC TAC		
240		
Lys Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr		
65	70	75
ATG CAA CTC AGC AGC CTG ACA TCT GAG GAC TCT GCC GTC TAT TAC TGT		
288		
Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys		
85	90	95
TCA AGG TCC TAC GAC TTT GCC TGG TTT GCT TAC TGG GGC CAA GGG ACT		
336		
Ser Arg Ser Tyr Asp Phe Ala Trp Phe Ala Tyr Trp Gly Gln Gly Thr		
100	105	110
CCG GTC ACT GTC TCT GCA		
354		
Pro Val Thr Val Ser Ala		
115		

## (2) INFORMATION FOR SEQ ID NO: 74:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 118 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Gln	Val	Gln	Leu	Gln	Gln	Ser	Gly	Ala	Glu	Leu	Met	Lys	Pro	Gly	Ala	
1				5				10			15					
Ser Val Lys Ile Ser Cys Lys Ala Thr Gly Tyr Thr Phe Ser Ala Tyr																
				20				25			30					
Trp Ile Glu Trp Val Lys Gln Arg Pro Gly His Gly Leu Glu Trp Ile																
				35				40			45					
Gly Glu Ile Leu Pro Gly Ser Asn Asn Ser Arg Tyr Asn Glu Lys Phe																
				50				55			60					
Lys Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Asn Thr Ala Tyr																
				65				70			75			80		
Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys																
				85				90			95					
Ser Arg Ser Tyr Asp Phe Ala Trp Phe Ala Tyr Trp Gly Gln Gly Thr																
				100				105			110					
Pro Val Thr Val Ser Ala																
				115												

## (2) INFORMATION FOR SEQ ID NO: 75:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 342 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mouse

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..342

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GAC ATT GTG ATG TCA CAG TCT CCA TCC TCC CTA GCT GTG TCA GTT GGA  
 48  
 Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Ala Val Ser Val Gly  
 1 5 10 15  
 GAG ARG GTT ACT ATG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT  
 96  
 Glu Lys Val Thr Met Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser  
 20 25 30  
 AGC AAT CRA AAG ATC TAC TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG  
 144  
 Ser Asn Gln Lys Ile Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln  
 35 40 45  
 TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC ACT AGG GAA TCT GGG GTC  
 192  
 Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Thr Arg Glu Ser Gly Val  
 50 55 60  
 CCT GAT CGC TTC ACA GGC GGT GGA TCT GGG ACA GAT TTC ACT CTC ACC  
 240  
 Pro Asp Arg Phe Thr Gly Gly Ser Gly Thr Asp Phe Thr Leu Thr  
 65 70 75 80  
 ATC AGC AGT GTG AAG GCT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAA  
 288  
 Ile Ser Ser Val Lys Ala Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln  
 85 90 95  
 TAT TAT AGA TAT CCT CGG ACG TTC GGT GGA GGC ACC AAG CTG GAA ATC  
 336  
 Tyr Tyr Arg Tyr Pro Arg Thr Phe Gly Gly Thr Lys Leu Glu Ile  
 100 105 110  
 AAA CGG  
 342  
 Lys Arg

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

(B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

Asp	Ile	Val	Met	Ser	Gln	Ser	Pro	Ser	Ser	Leu	Ala	Val	Ser	Val	Gly
1					5				10				15		
Glu	Lys	Val	Thr	Met	Ser	Cys	Lys	Ser	Ser	Gln	Ser	Leu	Leu	Tyr	Ser
			20					25				30			
Ser	Asn	Gln	Lys	Ile	Tyr	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Gln
			35					40			45				
Ser	Pro	Lys	Leu	Leu	Ile	Tyr	Trp	Ala	Ser	Thr	Arg	Glu	Ser	Gly	Val
	50						55			60					
Pro	Asp	Arg	Phe	Thr	Gly	Gly	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr
	65				70				75		80				
Ile	Ser	Ser	Val	Lys	Ala	Glu	Asp	Leu	Ala	Val	Tyr	Tyr	Cys	Gln	Gln
			85					90				95			
Tyr	Tyr	Arg	Tyr	Pro	Arg	Thr	Phe	Gly	Gly	Gly	Thr	Lys	Leu	Glu	Ile
			100				105					110			
Lys	Arg														

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 354 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS  
 (B) LOCATION: 1..354

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

CAG	GTG	CAG	CTG	GTG	CAG	TCT	GGG	GCA	GAG	GTG	AAA	AAG	CCT	GGG	GCC		
						48											
Gln		Gln		Leu		Gln		Ser	Gly	Ala	Glu	Val	Lys	Lys	Pro	Gly	Ala
1						5					10			15			
TCA	GTG	AAG	GTG	TCC	AAG	GCT	TCT	GCC	TAC	ACC	TTC	AGT	GCC	TAC			
							96										
Ser	Val	Lys	Val	Ser	Cys	Lys	Ala	Ser	Gly	Tyr	Thr	Phe	Ser	Ala	Tyr		
						20				25		30					
TGG	ATA	GAG	TGG	GTG	CGC	CAG	GCT	CCA	GGG	AAG	GCC	CTC	GAG	TGG	GTC		
						144											
Trp	Ile	Glu	Trp	Val	Arg	Gln	Ala	Pro	Gly	Lys	Gly	Leu	Glu	Trp	Val		

74

35

40

45

GGA GAG ATT TTA CCT GGA AGT AAT AAT TCT AGA TAC AAT GAG AAG TTC  
 192  
 Gly Glu Ile Leu Pro Gly Ser Asn Asn Ser Arg Tyr Asn Glu Lys Phe  
 50 55 60

AAG GGC CGA GTG ACA GTC ACT AGA GAC ACA TCC ACA AAC ACA GCC TAC  
 240  
 Lys Gly Arg Val Thr Val Thr Arg Asp Thr Ser Thr Asn Thr Ala Tyr  
 65 70 75 80

ATG GAG CTC AGC AGC CTC AGG AGG TCT GAG GAC ACA GCC GTC TAT TAC TGT  
 288  
 Met Glu Leu Ser Ser Leu Arg Ser Glu Asp Thr Ala Val Tyr Tyr Cys  
 85 90 95

GCA AGA TCC TAC GAC TTT GCC TGG TTT GCT TAC TGG GGC CAA GGG ACT  
 336  
 Ala Arg Ser Tyr Asp Phe Ala Trp Phe Ala Tyr Trp Gly Gln Gly Thr  
 100 105 110

CTG GTC ACA GTC TCC TCA  
 354  
 Leu Val Thr Val Ser Ser  
 115

(2) INFORMATION FOR SEQ ID NO: 78:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 118 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Lys Lys Pro Gly Ala  
 1 5 10 15

Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Ser Ala Tyr  
 20 25 30

Trp Ile Glu Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val  
 35 40 45

Gly Glu Ile Leu Pro Gly Ser Asn Asn Ser Arg Tyr Asn Glu Lys Phe  
 50 55 60

Lys Gly Arg Val Thr Val Thr Arg Asp Thr Ser Thr Asn Thr Ala Tyr  
 65 70 75 80

Met Glu Leu Ser Ser Leu Arg Ser Glu Asp Thr Ala Val Tyr Tyr Cys  
 85 90 95

Ala Arg Ser Tyr Asp Phe Ala Trp Phe Ala Tyr Trp Gly Gln Gly Thr  
 100 105 110

Leu Val Thr Val Ser Ser  
 115

(2) INFORMATION FOR SEQ ID NO: 79:

- (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 342 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS  
 (B) LOCATION: 1..342

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GAC ATC CAG ATG ACC CAG AGC CCA AGC AGC CTG AGC GCC AGC GTG GGT  
 48  
 Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly  
 1 5 10 15

GAC AGA GTG ACC ATC ACC TGT AAG TCC AGT CAG AGC CTT TTA TAT AGT  
 96  
 Asp Arg Val Thr Ile Thr Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser  
 20 25 30

AGC AAT CAA AAG ATC TAC TTG GCC TGG TAC CAG CAG AAG CCA GGT AAG  
 144  
 Ser Asn Gln Lys Ile Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys  
 35 40 45

GCT CCA AAG CTG CTG ATC TAC TGG GCA TCC ACT AGG GAA TCT GGT CTG  
 192  
 Ala Pro Lys Leu Leu Ile Tyr Trp Ala Ser Thr Arg Glu Ser Gly Val  
 50 55 60

CCA AGC AGA TTC AGC GGT AGC GGT ACC GAC TTC ACC ACC  
 240  
 Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Phe Thr  
 65 70 75 80

ATC AGC CTC CAG CCA GAG GAC ATC GCC ACC TAC TAC TGC CAG CAA  
 288  
 Ile Ser Ser Leu Gln Pro Glu Asp Ile Ala Thr Tyr Tyr Cys Gln Gln  
 85 90 95

TAT TAT AGA TAT CCT CGG ACC TTC GGC CAA GGG ACC AAG GTG GAA ATC  
 336  
 Tyr Tyr Arg Tyr Pro Arg Thr Phe Gly Gln Gly Thr Lys Val Glu Ile  
 100 105 110

AAA CCT  
 342  
 Lys Arg

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 114 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly  
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser  
20 25 30

Ser Asn Gln Lys Ile Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys  
35 40 45

Ala Pro Lys Leu Leu Ile Tyr Trp Ala Ser Thr Arg Glu Ser Gly Val  
50 55 60

Pro Ser Arg Phe Ser Gly Ser Gly Thr Asp Phe Thr Phe Thr  
65 70 75 80

Ile Ser Ser Leu Gln Pro Glu Asp Ile Ala Thr Tyr Tyr Cys Gln Gln  
85 90 95

Tyr Tyr Arg Tyr Pro Arg Thr Phe Gly Gln Gly Thr Lys Val Glu Ile  
100 105 110

Lys Arg

CLAIMS

1. A virus, or virus-like particle, derived from a virus or virus-like particle having a receptor for a host cell comprising a modified binding specificity conferred by a binding moiety allowing the virus or virus-like particle to bind to a target cell characterised in that the said host cell receptor is modified or absent so that the virus or virus-like particle is substantially incapable of binding the said host cell.
- 10 2. A virus or virus-like particle according to Claim 1 wherein the target cell is eukaryotic.
- 15 3. A virus or virus-like particle according to Claim 2 that is an adenovirus, influenza virus, vaccinia virus, any other animal virus or replication-defective derivative of any of these.
- 20 4. An adenovirus or influenza virus or vaccinia virus, or a replication defective derivative of any of these, characterized in that the virus has a modified binding specificity conferred by a binding moiety allowing the virus to bind to a target cell.
- 25 5. A virus or virus-like particle according to any of Claims 1 to 4 wherein the binding moiety is a monoclonal antibody, an ScFv, a dAb, or a minimal recognition unit of an antibody.
- 30 6. A virus or virus-like particle according to any of Claims 1 to 4 wherein the binding moiety is at least part of a ligand of a target cell-specific cell-surface receptor.

7. A virus or virus-like particle according to Claim 5 or 6 wherein the binding moiety recognises a target cell-specific surface antigen.
- 5 8. A virus according to any one of Claims 1 to 7 wherein the binding moiety is joined to a molecule on the virus or virus-like particle other than the receptor for its host cell.
- 10 9. A virus or virus-like particle according to any one of Claims 1 to 7 wherein the binding moiety is joined to or forms part of the receptor on the said virus or virus-like particle for its host.
- 15 10. A virus or virus-like particle according to Claim 8 wherein the said molecule on the surface of the virus or virus-like particle is a protein.
11. A virus or virus-like particle according to Claim 6 wherein the target cell-specific cell-surface receptor is any one of GnRH receptor, MSH receptor and somatostatin receptor.
- 20 12. A virus or virus-like particle according to any one of Claims 1 to 11 modified further to contain a gene suitable for gene therapy.
- 25 13. A virus or virus-like particle according to Claim 12 wherein the gene encodes a molecule having a directly or indirectly cytotoxic function.
14. A virus or virus-like particle according to Claim 13 wherein the gene encodes any one of interleukin-2, tumour necrosis factor, interferon-gamma, ribonuclease and deoxyribonuclease.

15. A virus or virus-like particle according to Claim 13 wherein the gene encodes an enzyme capable of converting a relatively non-toxic pro-drug into a cytotoxic drug.
- 5 16. A virus or virus-like particle according to Claim 15 wherein the gene is either cytosine deaminase or thymidine kinase.
17. A virus or virus-like particle according to Claim 12 wherein the gene overcomes a defect in a gene in the target cell.
- 10 18. A virus or virus-like particle according to Claim 17 wherein the gene is any one of CFTR, dystrophin and haemoglobin A.
- 15 19. A virus, or virus-like particle, containing nucleic acid, according to any one of Claims 1 to 15 wherein the said virus or virus-like particle is adapted to deliver the said nucleic acid to the target cell.
20. A virus or virus-like particle according to Claim 1 wherein the said receptor comprises protein.
21. A virus according to Claim 20 wherein the virus is influenza virus and the said receptor is the haemagglutinin receptor protein.
- 25 22. A virus according to Claim 20 wherein the virus is adenovirus and the said receptor is the penton fibre protein.
23. A virus according to Claim 22 wherein the binding moiety is fused to the penton fibre protein at any one or more of the junctions of the repetitive units of the shaft.
- 30

24. A virus according to Claim 23 wherein the binding moiety is a ScFv.
25. A virus according to Claim 24 wherein the ScFv binds to a 5 tumour cell antigen.
26. A virus or virus-like particle according to any one of Claims 1 to 25 wherein the binding moiety is a polypeptide.
- 10 27. A virus or virus-like particle according to Claim 26 when dependent on either of Claims 10 or 20 wherein the binding moiety is fused to the protein on the surface of the said virus or virus-like particle.
- 15 28. A virus or virus-like particle according to any one of Claims 1 to 27 for use in medicine.
29. A nucleotide sequence encoding the fusion of the binding moiety and the protein on the surface of the virus or virus-like particle 20 according to any one of Claims 23 to 25 and 27.
30. A nucleotide sequence encoding the receptor modified as defined in Claim 8, wherein the receptor comprises a polypeptide backbone.
- 25 31. A nucleotide sequence defined in any of Claims 29 or 30 additionally comprising the remainder of the genome of the virus or virus-like particle.
- 30 32. A nucleotide sequence encoding a virus or virus-like particle

according to any one of Claims 1 to 27.

33. A therapeutic system comprising a virus or virus-like particle according to Claim 15 or 16 and a pro-drug.

5

34. A method for producing a virus or virus-like particle according to any of Claims 1 to 27 in cell culture, the method comprising (1) infecting the cells with the said virus or virus-like particle, (2) culturing the infected cells until the virus or virus-like particle reaches a sufficiently high titre, (3) harvesting and substantially purifying the virus or virus-like particle and (4) joining the binding moiety to the substantially purified virus or virus-like particle.

10

15 35. A method for producing a virus or virus-like particle according to any of Claims 1 to 27 in cell culture, the method comprising (1) genetically modifying the virus or virus-like particle to produce a binding moiety, (2) infecting cells with the genetically modified virus or virus-like particle, (3) culturing the cells until the virus or virus-like particle reaches a sufficiently high titre and (4) harvesting and substantially purifying the genetically modified virus or virus-like particle.

20

25 36. A pharmaceutical composition comprising a virus or virus-like particle according to any one of Claims 1 to 27 and a pharmaceutical carrier.

30

37. A method of treating a mammal having target cells to be destroyed, the method comprising administering the virus or virus-like particle according to Claim 13.

38. A method of treating a mammal having target cells to be destroyed, the method comprising (1) administering a virus or virus-like particle according to Claim 15 or 16, (2) allowing the virus or virus-like particle to bind to and deliver its nucleic acid to the target cell and (3) administering the said pro-drug.

5

39. A method of treating a mammal having a defective gene, the method comprising administering the virus or virus-like particle according to Claim 17 or 18.

1/23

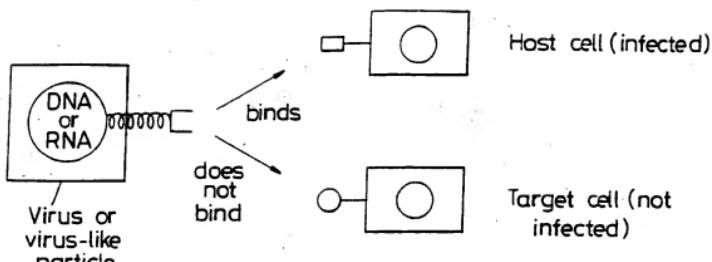


Fig. 1(a)

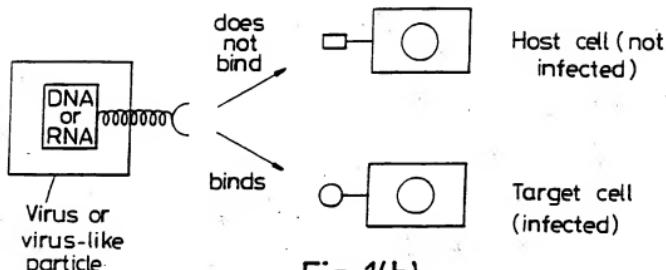


Fig. 1(b)

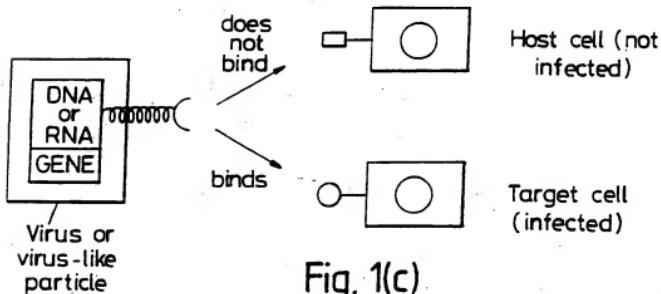


Fig. 1(c)

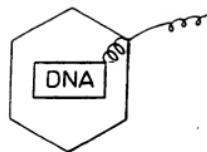


Fig. 2(a)

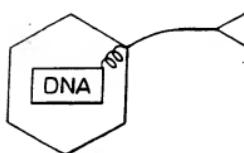


Fig. 2(b)

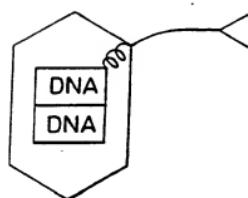


Fig. 2(c)

3/23

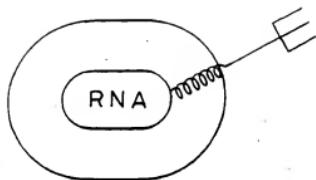


Fig. 3(a)

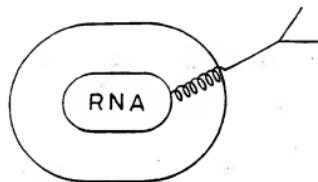


Fig. 3(b)

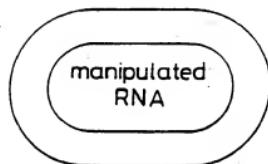


Fig. 4(a)

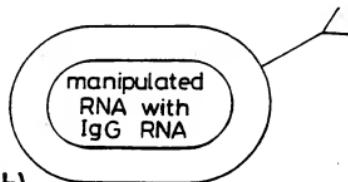


Fig. 4(b)

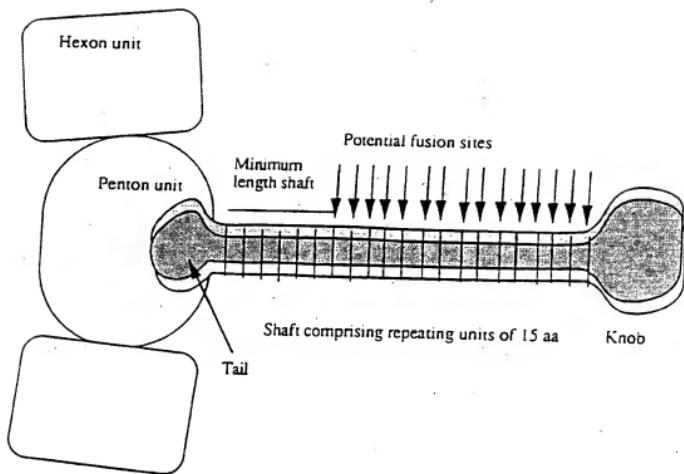


Figure 5

5/23

## Fusion A

	1	2	3	4	108	109
P L V T S N V Q L Q .....	L	E	*	*		
CCTCTAGTTACCTCCAATGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31210		PstI	..		XbaI	EcoRI

## Fusion B

	1	2	3	4	108	109
L S L D E A V Q L Q .....	L	E	*	*		
CTCTCTGGACGAGGCCGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31260		PstI	..		XbaI	EcoRI

## Fusion C

	1	2	3	4	108	109
P L K K T K V Q L Q .....	L	E	*	*		
CCTCTCAAAAAACCAAGGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31310	31320		PstI	..	XbaI	EcoRI

## Fusion D

	1	2	3	4	108	109
P L T V T S V Q L Q .....	L	E	*	*		
CCCCTCACAGTTACCTCAGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31360		PstI	..		XbaI	EcoRI

## Fusion E

	1	2	3	4	108	109
P L M V A D V Q L Q .....	L	E	*	*		
CCTCTAATGGTCGGGGCGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31400	31410		PstI	..	XbaI	EcoRI

Figure 6 (Page 1 of 5)

6/23

## Fusion F

	1	2	3	4	108	109									
P	L	T	V	H	D	V	Q	L	Q	.....	L	E	*	*	
CCGCTAACCGTGCACGACGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>															
31450							PstI ..			.XhoI			EcoRI		

## Fusion G

	1	2	3	4	108	109								
P	L	T	V	S	E	V	Q	L	Q	.....	L	E	*	*
CCCTCACAGTGTGCAGAACGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31490	31500				PstI ..			.XhoI			EcoRI			

## Fusion H

	1	2	3	4	108	109								
L	T	T	T	D	S	V	Q	L	Q	.....	L	E	*	*
CTCACACCACCGATAGCGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31540	31550				PstI ..			.XhoI			EcoRI			

## Fusion I

	1	2	3	4	108	109									
P	L	T	T	A	T	V	Q	L	Q	.....	L	E	*	*	
CCTCTAACTACTGCCACTGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>															
31590							PstI ..			.XhoI			EcoRI		

## Fusion J

	1	2	3	4	108	109								
P	I	Y	T	Q	N	V	Q	L	Q	.....	L	E	*	*
CCCATTATACACAAAATGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31630	31640				PstI ..			.XhoI			EcoRI			

Figure 6 (Page 2 of 5)

7/23

## Fusion K

	1	2	3	4	108	109								
H	V	T	D	D	L	V	Q	L	Q	.....	L	E	*	*
CATGTAACAGACGACCTAGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31730				PstI	..	.XhoI	EcoRI							

## Fusion L

	1	2	3	4	108	109								
G	V	T	I	N	N	V	Q	L	Q	.....	L	E	*	*
GGTGTGACTATTATAATGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31730				PstI	..	.XhoI	EcoRI							

## Fusion M

	1	2	3	4	108	109								
G	F	D	S	Q	G	V	Q	L	Q	.....	L	E	*	*
GGTTTGATTCACAAAGGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31780				PstI	..	.XhoI	EcoRI							

## Fusion N

	1	2	3	4	108	109								
R	I	D	S	Q	N	V	Q	L	Q	.....	L	E	*	*
AGGATTGATTCTCAAAACGTCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31830				PstI	..	.XhoI	EcoRI							

## Fusion O

	1	2	3	4	108	109								
F	D	A	Q	N	Q	V	Q	L	Q	.....	L	E	*	*
TTTGATGCTCAAAACCAAGTGCAGCTGCAG...ScFv... <u>CTCGAGTAATAAGAATT</u>														
31880				PstI	..	.XhoI	EcoRI							

Figure 6 (Page 3 of 5)

8/23

## Fusion P

	1	2	3	4	108	109
P F I N S A V Q L Q .....	L	E	*	*		
CTTTTTATAAACTCAGCCGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
31920					PstI ..	XbaI
						EcoRI

## Fusion Q

	1	2	3	4	108	109
S N N S K N V Q L Q .....	L	E	*	*		
TCAAACAATTCCAAAAACGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
32030	320401				PstI ..	XbaI
						EcoRI

## Fusion R

	1	2	3	4	108	109
G L M F D A V Q L Q .....	L	E	*	*		
GGGTTGATGTTGACGCTGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
32030	320401				PstI ..	XbaI
						EcoRI

## Fusion S

	1	2	3	4	108	109
P N A P N T V Q L Q .....	L	E	*	*		
CCTAAATGCACCAACACAGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
32100					PstI ..	XbaI
						EcoRI

## Fusion T

	1	2	3	4	108	109
L E F D S N V Q L Q .....	L	E	*	*		
CTAGAATTGATTCAAACAGTGCAG <u>CTGCAG</u> ...ScFv... <u>CTCGAGTAATAAGAATT</u> C						
32150					PstI ..	XbaI
						EcoRI

Figure 6 (Page 4 of 5)

9/23

## Fusion U

	1	2	3	4		108	109							
L	S	F	D	S	T	V	Q	L	Q	.....	L	E	*	*
CTTAGTTTGACAGCACAGTGCAG <u>CTGCAG</u> ...	ScFv...	<u>CTCGAG</u> TAATAAGAATTC												
32190						PstI	..				XbaI			EcoRI

## Fusion V

	1	2	3	4		108	109							
I	D	K	L	T	L	V	Q	L	Q	.....	L	E	*	*
ATTGATAAGCTAACTTTGGTGCAG <u>CTGCAG</u> ...	ScFv...	<u>CTCGAG</u> TAATAAGAATTC												
32240						PstI	..				XbaI			EcoRI

Figure 6 (Page 5 of 5)

10/23

LEADHBACK      AGCTAAGCTTGCATGCAAATTC  
*Hind*III *Sphi*

LEADBFOR      *peB* leader:  
P A M A R S Q L Q  
CCAGCGATGCCAGATCTCAGCTGCAGAGCT  
*Bgl*III      *Pst*I

Figure 7

11/23

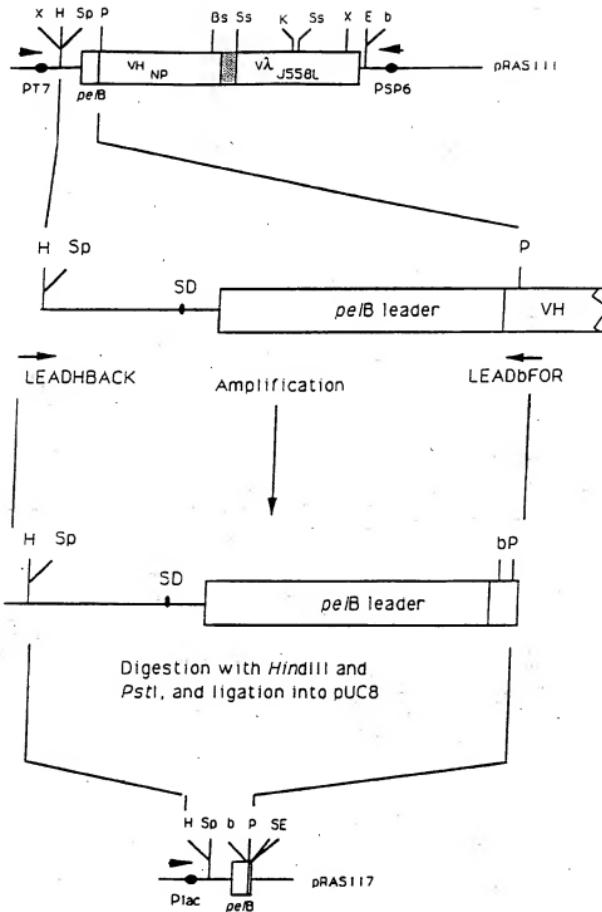


Figure 8

12/23

10 20 30 40 50

M K Y L L P

AAGCTTGCATGCAAATTCTATTCAGGAGACAGTCATAATGAAATACCTATTGCCT*HindIII SphI*

SD

/-----

60 70 80 90 100 110

T A A A G L L L L A A Q P A M A R S Q

ACGGCAGGCCGCTGGATTGTTATTACTCGCTGCCAACCCAGCGATGGCCAGATCTCAG-----peb leader-----| *BglII/V<sub>H</sub>*

120 130

L Q V D G S

CTGGCAGGTGGACGGGATCC*PstI SalI EcoRI*

Figure 9

13/23

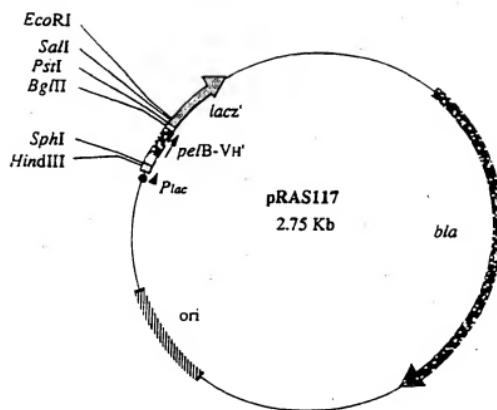


Figure 10

14/23

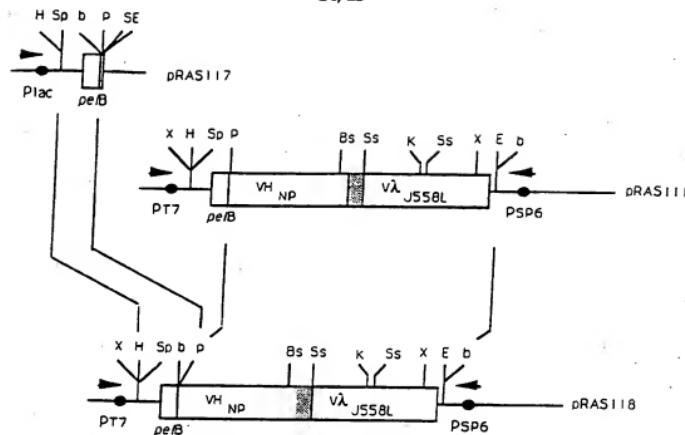


Figure 11

15/23

TAILbBACK

R S M K R A R P  
AGCTAGATCTATGAAGCCGC AAGACCG  
*BglII*

FIBRE3FOR

--fibre-----/--scFv-----  
 P L N R A R Q V Q L Q  
CCTCTCAAAAAACCAAGCAGGTGCAGCTGCAGCAGCCTGG  
*PstI*

FIBRE6FOR

--fibre-----/--scFv-----  
 P L T V H D Q V Q L Q  
CCCGCTAACCGTGCACGACCAGGTGCAGCTGCAGCAGCCTGG  
*PstI*

FIBRE9FOR

--fibre-----/--scFv-----  
 P L T T A T Q V Q L Q  
CCTCTAACTACTGCCACTCAGGTGCAGCTGCAGCAGCCTGG  
*PstI*

FIBRE12FOR

--fibre-----/--scFv-----  
 G V T I N N Q V Q L Q  
GGTGTGACTATTAATAATCAGGTGCAGCTGCAGGACCCCTGG  
*PstI*

FIBRE15FOR

--fibre-----/--scFv-----  
 P F D A Q N Q Q V Q L Q  
CCGTTTGATGCTAAACCAACAGGTGCAGCTGCAGCAGCC  
*PstI*

Figure 12 (Page 1 of 2)

16/23

FIBRE18FOR

--fibre-----/--scFv-----

G L M F D G Q V Q L Q

GGGTTGATGTTGACGCTCAGGTGCAGCTGCAGCAGCC

PstI

FIBRE21FOR

--fibre-----/--scFv-----

L S F D S T Q V Q L Q

GCCTTAGTTTACAGCACACAGGTGCAGCTGCAGCAGCC

PstI

FIBRE22FOR

--fibre-----/--scFv-----

G N K N N D K L T L Q V Q L Q

GGAAACAAAAATAATGATAAGCTAACTTGCAGGTGCAGCTGCAGCC

PstI

FIBREPFOR

--fibre-----

Y I A Q E \*

CATACATTGCCAAGAATAACAGGTGCAGCTGCAGCAGCCTGG

PstI

Figure 12 (Page 2 of 2)

17/23

10 20 30 40 50  
 M K Y L L P

**HindIII Sphi**  
 SD /-----

60 70 80 90 100 110  
 T A A A G L L L A A O P A M A Q V Q  
 ACGGCAGCCGCTGGATTGTTATTACTCGCTGCCAACCGCGATGGCCAGGTGCAAG  
 -----peI B leader-----/VH<sub>NP</sub>

120 130 140 150 160 170  
 L Q Q P G A E L V K P G A S V K L S C  
 CTGCAGCAGCCTGGGGCTGAGCTTGTGAAGCCTGGGGCTCAGTGAAGCTGTCCCTGC  
 PstI

180 190 200 210 220 2  
 K A S G Y T F T S Y W M H W V K Q R P  
 AAGGCTCTGGCTACACCTCACCGAGCTACTGGATGCACTGGGTGAAGCAGAGCCCT  
 CDR1

30 240 250 260 270 280  
 G R G L E W I G R I D P N S G G T R Y  
 GGACGAGGCCCTGAGTGGATTGGAAGGATGACTCTGGGTGACTAAGTAC  
 CDR2

290 300 310 320 330 340  
 N E K F L S K A T L T V D K P S S T A  
ATGAGAAAGTTCAGAGCAAGGCCACACTGACTGTAGACAACCCCTCCAGCACAGCC

350 360 370 380 390 40  
 Y M Q L S S L T S E D S A V Y Y C A R  
 TACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGGGTCTATTATCTGCAAGA

0 410 420 430 440 450  
 Y D Y Y G S S S Y F D Y W G Q G T T L T  
TACGATTACTACGGTAGTAGCTACTTGACTACTGGGCCAAGGGACCAAGGTCAACC  
 CDR3 BstEII

460 470 480 490 500 510  
 V S S G G G G S G G G G S G G G G S Q  
 GTCTCCTCAGGTGGAGGGCGGTTCAAGGGCGGAGGTGGCTCTGGCGGTGGCGGATCCAG  
 /-----(G<sub>4</sub>S)<sub>3</sub> linker-----/BamHI - /λ

520 530 540 550 560 570  
 V V L T Q E S A L T T S P G E T V T L  
 GCTGTTGTGACTCAGGAATCTGCACTCACCATCACCTGGTGAANCAAGTCAGCAACTC

580 590 600 610 620  
 T C R S S T G A V T T S N Y A N W V O  
 ACTTGTGGCTCAAGTACTGGGGCTTGTAAACTACTAATCTGCAACTGGGTCCAA  
 CDR1

Figure 13 (Page 1 of 2)

18/23

630	640	650	660	670	680
E K P D H L F T G L I G G T N N N R A P					
GAAAAAACAGATCATTTATTCACTGGTCTAAATAGGTGGTACCAAAACAAACCCAGCTCCA					
				Kpn1	CDR2 Sst1
690	700	710	720	730	740
G V P A R F S G S L I G D K A A L T I					
GGTGTTCCTGCCAGATTCTCAGGCTCCCTGATTGGAGACAAGGCTGCCCTCACCATC					
750	760	770	780	790	8
T G A Q T E D E A I Y F C A L W Y S N					
ACAGGGGACACAGACTGAGGATGAGGCAATATATTTCTGTGCTCTATGCTACAGC AAC					
				CDR3	
00	810	820	830	840	850
H W V F G G G T K L T V L G L E . .					
CACTGGGTGTTCCGTGGAGGAACCAAACTGACTGTCTAGGTCTCGAGTAATAAGAA					
				XbaI	ECO

TTC  
RI

Figure 13 (Page 2 of 2)

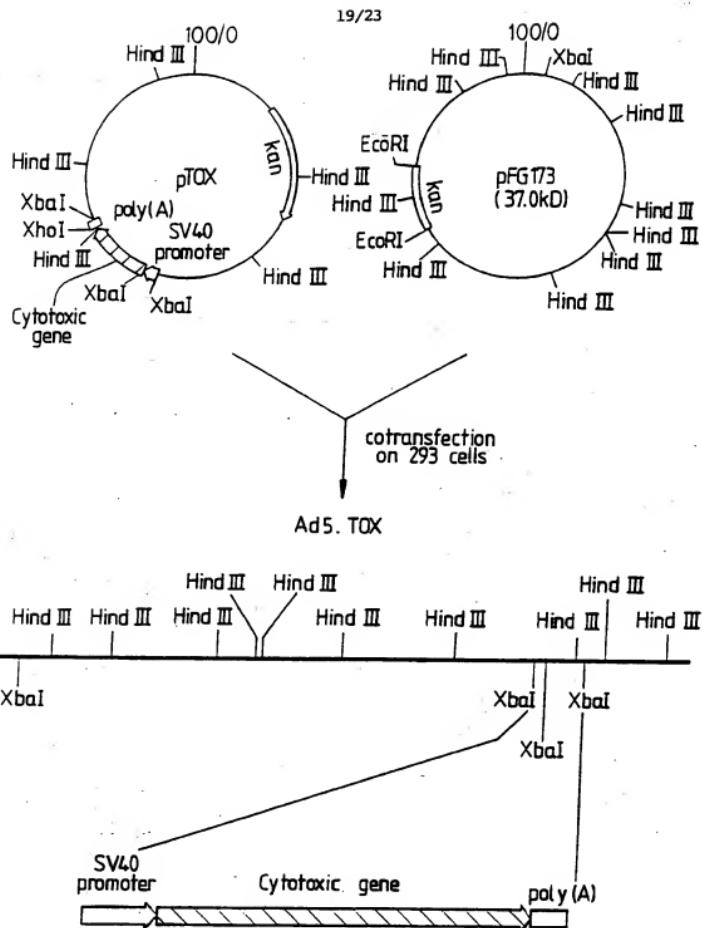


Fig. 14

(a)

MoVII - IIIFG1	Q	V	Q	L	Q	Q	S	C	A	E	L	H	K	P	G	A	S	V	K	15
MoVII - IIIFG1	CAGGTTCACCTGCCAGCTGGAGCTGAGCTGATGAAACCTGGGCCCCTAGTGAACTA																			
MoVII - IIIFG1	Q	V	O	L	Y	Q	S	G	A	E	Y	K	K	P	G	A	S	V	K	Y
	25	S	C	K	A	T	G	Y	T	F	S	Y	W	I	E	W	V	K	Q	R
MoVII - IIIFG1	TCCTGCCAGGGTACTGGCTAACATTCACTGGCCCTACTGGATAGTGCTAAAGCAGGC																			
MoVII - IIIFG1	S	C	K	A	S	G	Y	T	F	S	A	Y	W	I	E	W	V	B	Q	S
	35	CDR1	CDR2	40																
MoVII - IIIFG1	TCCTGCCAGGGTCTCGGCTAACCTCACTGGATAGTGCTAAAGCAGGC																			
MoVII - IIIFG1	P	G	H	G	L	E	W	I	G	E	I	L	P	G	S	N	N	S	R	Y
MoVII - IIIFG1	CCTGGACATGGCCCTGAGTGATTGGAGAGATTACCTGGAAAGTAATATTCTAGATAC																			
MoVII - IIIFG1	P	G	K	G	L	E	W	Y	G	E	I	L	P	G	S	N	N	S	R	Y

Figure 15 (Page 1 of 4)

21/23

Figure 15 (Page 2 of 4)

22/23

Figure 15 (Page 3 of 4)

23/23

<u>55</u> <u>E S</u> G V P D R F T G G C S G T D F T L T <u>HoVκ-HMEG1</u> GAAATCTGGGTCCCTGATGGCTCACAGGGGGATCTGGACAGATTCACTCTCACCC <u>HuVκ-HMEG1</u> GAAATCGTGTGCAAGGAGATTCAGGGTACCGGTAACGGTACCGTACCTTCACC	<u>60</u> <u>E S</u> G V P D R F T G G C S G T D F T L T <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA	<u>65</u> <u>E S</u> G V P D R F T G G C S G T D F T L T <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA	<u>70</u> <u>E S</u> G V P D R F T G G C S G T D F T L T <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA
<u>75</u> <u>E S</u> V K A E D L A V Y Y C Q Q Y Y R Y <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA	<u>80</u> <u>E S</u> V K A E D L A V Y Y C Q Q Y Y R Y <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA	<u>85</u> <u>E S</u> V K A E D L A V Y Y C Q Q Y Y R Y <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA	<u>90</u> (COR) <u>E S</u> V K A E D L A V Y Y C Q Q Y Y R Y <u>HoVκ-HMEG1</u> ATCAACAGTGTGAGGCTGAAGACCTGGCAGTTTACTGTAGCAAAATATAGATA <u>HuVκ-HMEG1</u> ATCAACAGCTCCAGCCAGGACATCGGCAACTACTGCCCAGCAATATAGATA
<u>95</u> <u>P R T</u> F C G G T K L E I K R <u>HoVκ-HMEG1</u> CCTCGGACCTTCTGGAGGCCAAAGCTGGAAATCAAAACCC <u>HuVκ-HMEG1</u> CCTCGGACCTTCTGGCCAGGGGACCAAGGTGGAAATCAAAACCT	<u>100</u> <u>P R T</u> F C G G T K L E I K R <u>HoVκ-HMEG1</u> CCTCGGACCTTCTGGAGGCCAAAGCTGGAAATCAAAACCC <u>HuVκ-HMEG1</u> CCTCGGACCTTCTGGCCAGGGGACCAAGGTGGAAATCAAAACCT	<u>105</u> <u>P R T</u> F C G G T K L E I K R <u>HoVκ-HMEG1</u> CCTCGGACCTTCTGGAGGCCAAAGCTGGAAATCAAAACCC <u>HuVκ-HMEG1</u> CCTCGGACCTTCTGGCCAGGGGACCAAGGTGGAAATCAAAACCT	

Figure 15 (Page 4 of 4)

## INTERNATIONAL SEARCH REPORT

Internat'l Application No  
PCT/US 93/02267

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 5 C12N15/87 A61K48/00 C12N15/34 C12N15/86 C12N7/04  
 C12N15/62 C12N15/13

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 5 C12N A61K C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	NUCLEIC ACIDS RESEARCH. vol. 21, no. 5, 11 March 1993, ARLINGTON, VIRGINIA US pages 1081 - 1085 RUSSELL, S.J. ET AL. 'Retroviral vectors displaying functional antibody fragments' cited in the application see the whole document	1-7, 9, 10, 12, 13, 15, 19, 20, 26-29, 31-38
X	HUMAN GENE THERAPY vol. 3, no. 2, April 1992 pages 147 - 154 CURIEL, D.T. ET AL. 'High-efficiency gene transfer mediated by adenovirus coupled to DNA-polylysine complexes' cited in the application see the whole document	1-3, 6, 8, 10, 12, 17, 19, 20, 26, 28, 33, 34, 36, 37

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "B" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another document of particular relevance (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or inventive if it is considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"A" document member of the same patent family

1 Date of the actual completion of the international search

Date of mailing of the international search report

7 February 1994

04-03-1994

Name and mailing address of the ISA  
 European Patent Office, P.B. 3818 Patentam 2  
 NL - 2280 HV Rijswijk  
 Tel. (+ 31-70) 340-2040, Tx. 31 651 cpo nl  
 Fax (+ 31-70) 340-3016

Authorized officer

Chambonnet, F

## INTERNATIONAL SEARCH REPORT

Internat Application No  
PCT/GB 93/02267

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TRENDS IN BIOTECHNOLOGY. vol. 9 , September 1991 , CAMBRIDGE GB pages 303 - 309 KINGSMAN, A.J. ET AL. 'Retroelement particles as purification, presentation and targeting vehicles' cited in the application see page 306, column 2, paragraph 2 - page 308, line 2	1,2,6, 8-10,12, 13,17, 19-21, 26,28, 31-33, 35-37,39
P,X	JOURNAL OF BIOLOGICAL CHEMISTRY. (MICROFILMS) vol. 268, no. 10 , 5 April 1993 , BALTIMORE, MD US pages 6866 - 6869 SHARON, I.M. ET AL. 'Binding-incompetent adenovirus facilitates molecular conjugate-mediated gene transfer by the receptor-mediated endocytosis pathway' cited in the application see the whole document	1-4
X	WO,A,92 06180 (UNIVERSITY OF CONNECTICUT) 16 April 1992	1,2, 6-10,12, 19,28, 31,32, 34,36
	see the whole document	
X	WO,A,92 14829 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 3 September 1992	1,2,6,7, 9,10,12, 13,17, 19,20, 26,28, 31,32, 35-38
	see the whole document	
X	EP,A,0 508 809 (BRITISH BIOTECHNOLOGY LIMITED) 14 October 1992	1,2, 6-10,12, 13,15, 19-21, 26-29, 31,35-38
	see the whole document	
P,X	WO,A,93 09221 (THERAGENE HB) 13 May 1993	1-5,7, 10,12, 13, 19-21, 26,28, 31,34-37
	see the whole document	

1

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 93/02267

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9206180	16-04-92	AU-A- 8860391 CA-A- 2092323 EP-A- 0553235	28-04-92 02-04-92 04-08-93
WO-A-9214829	03-09-92	AU-A- 8430291 EP-A- 0572401	15-09-92 08-12-93
EP-A-0508809	14-10-92	AU-A- 1538692 WO-A- 9218621	17-11-92 29-10-92
WO-A-9309221	13-05-93	SE-A- 9103183	01-05-93